Bias-Corrected Quantile Regression Estimation of Censored Regression Models

Cizek, P.; Sadikoglu, S.

Publication date:
2014

Document Version
Early version, also known as pre-print

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Cizek, P., & Sadikoglu, S. (2014). *Bias-Corrected Quantile Regression Estimation of Censored Regression Models*. (CentER Discussion Paper; Vol. 2014-060). Econometrics.

CentER

Discussion Paper

No. 2014-060

BIAS-CORRECTED QUANTILE REGRESSION ESTIMATION OF CENSORED REGRESSION MODELS

By

P. Čížek, S. Sadikoglu

6 October, 2014

ISSN 0924-7815
ISSN 2213-9532

Tilburg University

Bias-corrected quantile regression estimation of censored regression models

P. Čížek[†,*], S. Sadikoglu[‡,**]

[†] *CentER, Department of Econometrics & OR, Tilburg University, The Netherlands*

[‡] *Department of Econometrics & OR, Tilburg University, The Netherlands*

Abstract

Motivated by weak small-sample performance of the censored regression quantile estimator proposed by Powell (1986a), two- and three-step estimation methods were introduced for estimation of the censored regression model under conditional quantile restriction. While those stepwise estimators have been proven to be consistent and asymptotically normal, their finite sample performance greatly depends on the specification of an initial estimator that selects the subsample to be used in subsequent steps. In this paper, an alternative semiparametric estimator is introduced that does not involve a selection procedure in the first step. The proposed estimator is based on the indirect inference principle and is shown to be consistent and asymptotically normal under appropriate regularity conditions. Its performance is demonstrated and compared to existing methods by means of Monte Carlo simulations.

JEL classification: C21, C24

Keywords: asymptotic normality, censored regression, indirect inference, quantile regression

1. Introduction

The (Type 1) censored regression model has been studied and extensively used in a wide range of applied economics literature. To estimate the parameters of censored regression models, the maximum likelihood estimator (MLE) is usually used under the assumption that the underlying errors have a distribution function with a known parametric form (e.g., that the error terms are normally distributed). Contrary to linear regression, the resulting estimator is sensitive to departures from the parametric assumptions about the error term distribution. If the employed assumptions do not hold, the MLE estimates are in general inconsistent (cf. Arabmazar and Schmidt, 1982).

[*] Corresponding author. Tel: +31 13 4668723. Fax: +31 13 4668320. Email: P.Cizek@uvt.nl.

[**] Email: S.Sadikoglu@uvt.nl.

To relax the strong assumptions of MLE, several semiparametric estimators have been introduced in the econometric literature. Relying on very weak identification assumptions, Powell (1984, 1986a) proposed the censored least absolute deviation (CLAD) and censored regression quantile (CRQ) estimators by imposing the restriction that the conditional quantile of the error term is zero. These consistent and asymptotically normal estimators were applied in many contexts (e.g., Fahr, 2004; Melenberg and van Soest, 1996), and furthermore, have been extended in many directions, which include random censoring (Honore et al., 2002; Portnoy, 2003) as well as panel-data models (Honore, 1992; Campbell and Honore, 1993). In practice, the CRQ estimator is very appealing due to its robustness to misspecification of the error-term distribution and of the form of heteroskedasticity. On the other hand, it is difficult to compute exactly since its objective function is non-differentiable and non-convex. The algorithm for the exact computation of CLAD and CQR was proposed by Fitzenberger (1997b), but its demands for computational time make it infeasible in applications involving many regressors. Nevertheless, this algorithm was used by Fitzenberg (1997a) in simple Monte Carlo experiments that demonstrated a more severe drawback of the CRQ estimator in small samples than the mean-biasedness and inefficiency documented by Paarsch (1984) and Moon (1989): the CRQ estimator exhibits a very heavy-tailed distribution in small samples. Note that these unfavorable finite-sample properties are shared to some extent also by some alternatives to CLAD such as the symmetrically censored least squares of Powell (1986b). Other alternative estimator such as those by Horowitz (1986) and Honore and Powell (1994) do not exhibit such heavy-tailed finite-sample distributions, but require the error terms and the explanatory variables being independent, which is a rather strong assumption.

Subsequently, Khan and Powell (2001) highlighted the inherent property of CLAD and CRQ that causes their poor small-sample performance: the joint identification of observations entering the objective function and of the quantile regression line. This gave rise to stepwise estimation procedures, for example, by Khan and Powell (2001) and Chernozkukov and Hong (2002). These methods select first a subset of observations to identify the quantile regression line and then apply the quantile regression (QR) on the selected observations. The first step can be achieved, for example, by a nonparametric selection procedure as in Buchinsky and Hahn (1998) and Khan and Powell (2001). Although asymptotically equivalent to an 'oracle' QR estimator, the selection procedure in the first step works at a cost in finite samples. Alternatively, Chernozhukov and Hong (2002) developed a three-step estimation method that involves a parametric first step to circumvent the "curse-of-dimensionality" problem posed by nonparametric selection procedures. Its performance in small samples does not however improve upon the two-step estimators in simple regression models.

As the finite-sample behavior of the stepwise estimators does not seem substantially better than CLAD in studies of Khan and Powell (2001), Chernozhukov and Hong (2002), and most recently Tang et al. (2011), we introduce an alternative semiparametric estimator for censored regression model under conditional quantile restriction. Contrary to the existing methods, we apply the linear regression QR estimator to all data (rather than to a preselected subsample) and then correct its bias caused by censoring. For the bias correction, indirect inference (II), which was suggested by Gourieroux and Monfort (1993), is used. The indirect inference methodology is a simulation-based technique that is essentially used for estimation of the parameters of correctly specified but intractable models, but it that can be employed as a bias correction method too (e.g., Gourieroux et al., 2000, and Gourieroux et al., 2010).

Implementing the standard II approach requires knowledge of the error-term distribution at least up to a parametric form. To exploit only the conditional quantile restriction, we propose a new methodology to simulate values of the error terms from a semiparametrically estimated distribution. The proposed II estimator is based on the standard linear QR for two reasons. First, linear QR has desirable properties such as convexity of the objective function and a reasonably small variance in small samples. Second and more importantly, the properties of linear QR are known even under model misspecification (see Angrist et al., 2006) and can be used to construct a nonparametrically estimated error distribution for the II simulations and subsequent bias correction. Hence, the proposed bias-corrected QR procedure can be shown to be consistent and asymptotically normal. Its benefits in small samples are demonstrated by means of Monte Carlo simulations.

The remainder of the paper is organized as follows. Section 2 presents a review of relevant estimation methods of censored regression model and a brief overview of the indirect inference methodology. In Section 3, the proposed bias-corrected QR estimator is described in details. The asymptotic properties of the indirect estimator are discussed in Section 4 and the results of Monte Carlo experiments are presented in Section 5. Proofs are given in the appendices.

2. Estimation of censored regression model and indirect inference

In this section, the censored regression model and some relevant estimators are introduced in Section 2.1 and the indirect inference concept is described in Section 2.2.

2.1. Censored regression model

Let us define the censored regression model. First, data are supposed to be a random sample of size $n \in N$ originating from a latent linear regression model

$$y_i^* = x_i^T \beta^0 + \varepsilon_i, \tag{1}$$

where $y_i^* \in R$ is the latent dependent variable, $x_i \in R^k$ is the vector of explanatory variables, β^0 represents the k-dimensional parameter vector, and ε_i is the unobserved error term with its conditional τ-quantile, $\tau \in (0, 1)$, being zero: $q_\tau(\varepsilon_i|x_i) = 0$. The observed responses y_i equal to y_i^* censored from below at some c_i:

$$y_i = \max\{c_i, x_i^T \beta^0 + \varepsilon_i\}. \tag{2}$$

We consider here only the case of fixed censoring with a known cut-off point $c_i \equiv c$, and without loss of generality, $c = 0$. An extension to random censoring is possible by the procedure of Honore et al. (2002).

The CRQ estimator is an extension of the classical linear QR to the censored regression model under a conditional quantile restriction. Since the conditional quantile function of y_i in (2) is simply $\max\{0, x_i^T \beta^0\}$, Powell (1986a) proposed the CRQ estimator $\widehat{\beta}_n^{CRQ}$:

$$\widehat{\beta}_n^{CRQ} = \arg\min_{\beta \in B} \sum_{i=1}^{n} \rho_\tau(y_i - \max\{0, x_i^T \beta\}), \tag{3}$$

where B is a compact parameter space, $\rho_\tau(z) = \{\tau - I\,(z \leq 0)\} \cdot z$ with $\tau \in (0, 1)$, and $I(\cdot)$ denotes the indicator function. Note that CRQ can be interpreted as applying the linear QR estimator to the observations x_i with $x_i^T \beta^0 \geq 0$ because the residuals of the observations with $x_i^T \beta^0 < 0$ do not carry any information about β^0. This leads then to a heavy-tailed small-sample distribution of CRQ.

To eliminate this property, Khan and Powell (2001) proposed two-step estimation method. In the first step, the observations with $x_i^T \beta^0 > 0$ are determined by an initial semiparametric or nonparametric estimation, and in the second step, the standard QR estimation is conducted on the selected observations. Nevertheless, the finite sample results of Khan and Powell (2001) do not seem to generate a substantial advantage with respect to the CRQ estimator in terms of mean or median squared errors, possibly due to an imprecise selection of observations in the first step; alternatively, using a local rather than a global optimization algorithm for CRQ could have played a role.

Later, Chernozhukov and Hong (2002) proposed a semiparametric three-step estimator of the censored regression model under conditional quantile restriction. The initial subset of observations with $x_i^T\beta^0 > 0$ is selected by a parametric binary-choice model (e.g., logit) and QR is used in the subsequent steps to obtain not only consistent estimates, but also a more precise selection of the observations with $x_i^T\beta^0 > 0$. Their finite-sample results are however not substantially better than those of the two-step procedures: while having a smaller mean bias in small samples, the three-step estimates often exhibit a larger mean squared errors (cf. Tang et al., 2011, and Section 5).

2.2. Parametric indirect inference

Our strategy for estimating the censored regression model will differ from the existing ones in that QR will be applied to all observations and its bias due to censoring will be corrected by means of the indirect inference (II). In this section, we therefore describe a general principle of (parametric) II introduced by Gourieroux and Monfort (1993) and discuss how II can be applied as a bias correction method following Gourieroux et al. (2000).

Consider a general model, for example, (1)–(2):

$$y_i = h(x_i, \varepsilon_i; \beta), \tag{4}$$

where y_i represents the response variable, $x_i \in R^k$ is the vector of explanatory variables with a distribution function $G^0(\cdot)$, $\beta^0 \in B \subset R^k$ is the parameter vector, and ε_i is the unobserved error term with a known conditional distribution function $F^0(\cdot|x_i)$ (a generalization to a nonparametrically estimated distribution function will follow in Section 3).

To implement II, an instrumental criterion, which is a function of the observations $\{y_i, x_i\}_{i=1}^n$ and of an auxiliary parameter vector $\theta \in \Theta \subset R^q$, $q \geq k$, has to be defined (e.g., linear QR applied to censored data). This criterion is minimized to estimate the auxiliary parameter vector:

$$\widehat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\{y_i\}_{i=1}^n, \theta)$$

(please note that the dependence on the explanatory variables $\{x_i\}_{i=1}^n$ is kept implicit as we do not consider simulating values of x_i, but work conditionally on observed $\{x_i\}_{i=1}^n$; see Gourieroux et al., 2000, for details).

The data-generating process is then fully determined by F^0 and β^0 and the instrumental criterion is assumed to converge asymptotically to a non-stochastic limit that has a unique minimum θ^0:

$$\theta^0 = \arg\min_{\theta \in \Theta} Q_\infty(F^0, \beta^0, \theta).$$

Evaluating it at any $F(\cdot|x_i)$ and β leads to the definition of the binding function $b(F, \beta)$:

$$b(F, \beta) = \arg\min_{\theta \in \Theta} Q_\infty(F, \beta, \theta), \tag{5}$$

which implies that $\theta^0 = b(F^0, \beta^0)$.

Under some regularity assumptions, $\widehat{\theta}_n$ is a consistent estimator of θ^0. Provided that $b(F^0, \beta)$ is known and one-to-one, a consistent estimate $\widetilde{\beta}_n$ of β^0 would be defined as $\widetilde{\beta}_n = b^{-1}(F^0, \widehat{\theta}_n)$ (F^0 is traditionally assumed to be fully known; auxiliary parameters of the error distribution have to be a part of the parameter vector θ). Since the binding function is often difficult to compute, Gourieroux and Monfort (1993) defined a simulation-based procedure to estimate the parameter β^0.

Let $\{\widetilde{\varepsilon}^1, ..., \widetilde{\varepsilon}^s\}$ be S sets of error terms, where $\widetilde{\varepsilon}^s = \{\widetilde{\varepsilon}_i^s\}_{i=1}^n, s = 1, ..., S$, are simulated from $F^0(\cdot|x_i)$, $\widetilde{\varepsilon}_i^s|x_i \sim F^0(\cdot|x_i)$. Then for any given β, one can generate S sets of simulated paths $\{\widetilde{y}^1(\beta), ..., \widetilde{y}^S(\beta)\}$ using model (4), where $\widetilde{y}^s(\beta) = \{\widetilde{y}_i^s(\beta)\}_{i=1}^n$ and $\widetilde{y}_i^s(\beta) = h(x_i, \widetilde{\varepsilon}_i^s; \beta)$ conditional on x_i for $s = 1, ..., S$. From these simulated samples, S auxiliary estimates can be computed:

$$\widetilde{\theta}_n^s(\beta) = \arg\min_{\theta \in \Theta} Q_n(\{\widetilde{y}_i^s(\beta)\}_{i=1}^n, \theta). \tag{6}$$

Under appropriate conditions, $\widetilde{\theta}_n^s(\beta)$ tends asymptotically to $b(F^0, \beta)$, which allows to define the indirect inference estimator in the following way:

$$\widehat{\beta}_n^{II} = \arg\min_{\beta \in B} \left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S} \widetilde{\theta}_n^s(\beta)\right]^T \Omega \left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S} \widetilde{\theta}_n^s(\beta)\right], \tag{7}$$

where Ω is a positive definite weighting matrix. As in GMM estimation, the choice of Ω does not affect the asymptotic distribution of the estimator if $\dim(\beta) = \dim(\theta)$ and its choice will thus be irrelevant. This estimator can be shown to be consistent and asymptotically normal.

To argue that II can be used as a bias correction technique, note that β can represent the parameter value in the original model (4) (e.g., censored regression (1)–(2)) and θ the parameter value of the auxiliary biased criterion (e.g., linear QR applied to censored data). The binding function $b(F^0, \beta)$ then maps the parameter values β to biased estimates θ and its inverse $\widehat{\beta}_n^{II} = b^{-1}(F^0, \widehat{\theta}_n)$ maps the biased estimates back to the parameters in the original model; see Gourieroux et al. (2000) for details.

3. Semiparametric indirect inference for censored regression

In this subsection, we introduce the semiparametric indirect estimation procedure to estimate the parameter vector of the censored regression model under conditional quantile restriction. As the linear quantile regression is used as an instrumental criterion and the distribution of ε_i is unknown, a crucial ingredient of the procedure is the behavior of QR under misspecification. Angrist et al. (2006) characterize the QR vector under misspecification as a minimizer of a weighted mean-squared approximation to the true conditional quantile function, assuming the almost-sure existence of the density of the dependent variable. As this result does not directly apply to the censored regression model, we modify their result to accommodate the censored regression model.

Let us first introduce necessary notation. The conditional quantile function of the dependent variable y_i is $\max\{0, x_i^T\beta^0\}$. For any quantile index $\tau \in (0,1)$, the QR vector is defined by:

$$\theta^0 = \arg\min_{\theta \in \Theta} E[\rho_\tau(y_i - x_i^T\theta)], \tag{8}$$

where $\rho_\tau(z) = \{\tau - I(z \leq 0)\} \cdot z$. Further, let $\Delta(x_i, \beta^0, \theta)$ denote the QR specification error, $\Delta(x_i, \beta^0, \theta) = x_i^T\theta - \max\{0, x_i^T\beta^0\}$, and the observed residual u_i with a conditional distribution $F_u(u|x_i)$ and a conditional density $f_u(u|x_i)$ be defined by $u_i = y_i - \max\{0, x_i^T\beta^0\}$.

Theorem 1. *Suppose that $E(y_i)$ and $E\|x_i\|^2$ are finite, θ^0 uniquely solves (8), and $P\{\Delta(x_i, \beta^0, \theta^0) = 0\} = 0$. Then, $\bar{\theta} = \theta^0$ uniquely solves the equation*

$$\bar{\theta} = \arg\min_{\theta \in \Theta} E\left[w(x_i, \beta^0, \bar{\theta}) \cdot \Delta^2(x_i, \beta^0, \theta)\right], \tag{9}$$

where

$$w(x_i, \beta^0, \bar{\theta}) = \begin{cases} \frac{F_u\{\Delta(x_i, \beta^0, \bar{\theta})|x_i\} - \tau}{2\Delta(x_i, \beta^0, \theta)} & \text{if } \Delta(x_i, \beta^0, \bar{\theta}) \neq 0 \\ w_0(x_i) & \text{if } \Delta(x_i, \beta^0, \bar{\theta}) = 0 \end{cases}$$

for any bounded function $w_0(x_i) : R^k \to R_0^+$.

Proof: See Appendix A. □

Theorem 1 states that the linear QR vector depends on the weighting function $w(x_i, \beta^0, \theta^0)$, which in turn is a function of the distribution function $F_u(\cdot|x_i)$. Thus, for any other distribution function $\widetilde{F}_{\widetilde{u}}(\cdot)$ such that $\widetilde{F}_{\widetilde{u}}\{\Delta(x_i,\beta^0,\theta^0)|x_i\} = F_u\{\Delta(x_i,\beta^0,\theta^0)|x_i\}$, the weighting function remains unchanged and the linear QR yields the same vector θ^0.

Next, we consider Theorem 1 in the censored regression model with errors $\varepsilon_i \sim F_\varepsilon(\cdot|x_i)$. First, note that $u_i = \max\{\varepsilon_i, -x_i^T\beta^0\}$ for $x_i^T\beta^0 > 0$ and $u_i = y_i$ for $x_i^T\beta^0 < 0$. Consider now different errors $\widetilde{\varepsilon}_i \sim \widetilde{F}_{\widetilde{\varepsilon}}(\cdot|x_i)$: one can set again $\widetilde{u}_i = \max\{\widetilde{\varepsilon}_i, -x_i^T\beta^0\}$ for $x_i^T\beta^0 > 0$ and $\widetilde{u}_i = \widetilde{y}_i = \max\{0, x_i^T\beta^0 + \widetilde{\varepsilon}_i\}$ for $x_i^T\beta^0 < 0$. We will show now that the distribution $\widetilde{\varepsilon}_i$ can be a normal one: $\widetilde{\varepsilon}_i \sim N(\mu_\tau\sigma(x_i;\beta^0), \sigma(x;\beta^0))$, where μ_τ is τth conditional quantile of the standard normal distribution $N(0,1)$ and $\sigma^2(x_i;\beta^0)$ denotes the conditional variance. Specifically, we find $\sigma(x_i;\beta^0)$ such that $\widetilde{F}_{\widetilde{u}}\{\Delta(x_i,\beta^0,\theta^0)|x_i\} = F_u\{\Delta(x_i,\beta^0,\theta^0)|x_i\}$ for any finite value of x_i. First note that $F_u\{\Delta(x_i,\beta^0,\theta^0)|x_i\} = F_y(x_i^T\theta^0|x_i)$ (and analogously for $\widetilde{u}_i$ and $\widetilde{y}_i$): $\sigma(x_i;\beta^0)$ has to be therefore chosen so that

$$F_y(x_i^T\theta^0|x_i) = \widetilde{F}_{\widetilde{y}}(x_i^T\theta^0|x_i). \tag{10}$$

The definition of $\sigma(x_i;\beta^0)$ is irrelevant if $x_i^T\theta^0 < 0$ as then $F_y(x_i^T\theta^0|x_i) = \widetilde{F}_{\widetilde{y}}(x_i^T\theta^0|x_i) = 0$. Ignoring the case of $F_y(x_i^T\theta^0|x_i) = \tau$, which will be dealt with later, (10) for $x_i^T\theta^0 \geq 0$ means

$$F_y(x_i^T\theta^0|x_i) = \Phi\left(\frac{x_i^T\theta^0 - [x_i^T\beta^0 + \mu_\tau\sigma(x_i;\beta^0)]}{\sigma(x_i;\beta^0)}\right) = \Phi_\tau\left(\frac{x_i^T\theta^0 - x_i^T\beta^0}{\sigma(x_i;\beta^0)}\right) \tag{11}$$

and

$$\sigma(x_i;\beta^0) = \frac{x_i^T\theta^0 - x_i^T\beta^0}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))}, \tag{12}$$

where Φ and Φ_τ are the distribution functions of $N(0,1)$ and $N(\mu_\tau,1)$, respectively (note that (12) leads to $\sigma(x_i,\beta^0) = 0$ for $x_i^T\theta^0 < 0$). Therefore, having $\widetilde{\varepsilon}_i \sim N(\mu_\tau\cdot\sigma(x_i;\beta^0), \sigma(x_i;\beta^0))$ with $\sigma(x_i,\beta^0)$ defined in (12) yields the same linear QR vector as the real data generated under $\varepsilon_i \sim F_\varepsilon(\cdot|x_i)$ and the biases of the linear QR estimates in the censored regression model (1)–(2) both with the original data distribution $\varepsilon_i \sim F_\varepsilon(\cdot|x_i)$ and with the data generated from $\widetilde{\varepsilon}_i \sim N(\mu_\tau\cdot\sigma(x_i;\beta^0), \sigma(x_i;\beta^0))$ will be equal.

If the bias correction of linear QR is to be performed by II, we can simulate the set of error terms $\{\widetilde{\varepsilon}^1, ..., \widetilde{\varepsilon}^S\}$ from $N(\mu_\tau\cdot\sigma(x_i;\beta), \sigma(x_i;\beta))$ instead of the original data distribution and calibrate over $\beta \in B$ (provided that $\sigma(x_i,\beta)$ is known). However, β is not identified in this case because equation (11) holds for any value β substituted for β^0 if definition (12) is used at that β. To achieve identification, β^0 in (12) has to be replaced by an initial estimate or the denominator in (12) also has to depend on β. We consider the latter strategy to achieve good performance even in very small samples. It is well known that the identification of the parameter vector in (1)–(2) under conditional quantile restriction relies on the observations with positive values of the index $x_i^T\beta^0 > 0$ (Powell, 1984, 1986a) since $F_y(x_i^T\beta^0|x_i) = \tau$ only if $x_i^T\beta^0 > 0$. We also exploit this fact and we define $\widetilde{\sigma}(x_i;\beta)$ for $x_i^T\theta^0 \geq 0$ as

$$\widetilde{\sigma}(x_i;\beta) = \begin{cases} \frac{x_i^T\theta^0 - x_i^T\beta}{\Phi_\tau^{-1}(F_{y_i}(x_i^T\theta^0|x_i))} & \text{if } x_i^T\beta \leq 0, \\ \frac{x_i^T\theta^0 - x_i^T\beta}{\Phi_\tau^{-1}(\min\{\max\{F_y(x_i^T\theta^0|x_i) - F_y(x_i^T\beta|x_i) + \tau, 0\}, 1\})} & \text{if } x_i^T\beta > 0; \end{cases} \tag{13}$$

($\widetilde{\sigma}(x_i;\beta) = 0$ for $x_i^T\theta^0 < 0$). Since $\widetilde{\sigma}(x_i;\beta) = \sigma(x_i;\beta)$ only if $\beta = \beta^0$, (11) using $\widetilde{\sigma}(x_i;\beta)$ will hold only at $\beta \equiv \beta^0$ and the parameter vector β can be identified (see Lemma 6 for details). Further, as (13) becomes indeterminate if $F_y(x_i^T\theta^0|x_i) = \tau$ or $F_y(x_i^T\theta^0|x_i) = F_y(x_i^T\beta|x_i)$, we replace the definition of $\widetilde{\sigma}(x_i;\beta)$ in such cases by the limit of (13) for $x_i^T\theta^0 \to 0$ or $x_i^T(\theta^0 - \beta) \to 0$ so that the variance function $\widetilde{\sigma}(x_i;\beta)$ is continuous in x_i:

$$\widetilde{\sigma}(x_i;\beta) \quad = \quad \frac{\phi_\tau\{\Phi_\tau^{-1}[F_y(x_i^T\theta^0|x_i)]\}}{f_y(x_i^T\theta^0|x_i)} \quad \text{if } F_y(x_i^T\theta^0|x_i) = 0 \text{ or } F_y(x_i^T\theta^0|x_i) = F_y(x_i^T\beta|x_i), \tag{14}$$

where $\phi_\tau(\cdot)$ is the density function of $\Phi_\tau(\cdot)$ (note that the limit is the same for both cases of (13)).

With the definition (13) and (14) of $\widetilde{\sigma}(x_i;\beta)$, which assumes knowledge of the true conditional distribution $F_y(x_i^T\theta^0|x_i)$ at $x_i^T\theta^0$, we can define the infeasible indirect inference (III) estimator $\widehat{\beta}_n^{III}$ by

$$\widehat{\beta}_n^{III} = \operatorname*{arg\,min}_{\beta\in B} \left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta)\right]^T \Omega \left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta)\right], \tag{15}$$

where $\widetilde{\theta}_n^s(\beta) = \operatorname*{arg\,min}_{\theta\in\Theta} \sum_{i=1}^{n} \rho_\tau(\widetilde{y}_i^s(\beta) - x_i^T\theta)$ using simulated data $\widetilde{y}_i^s(\beta) = \max\{0, x_i^T\beta + \widetilde{\varepsilon}_i^s\}$, $\widetilde{\varepsilon}_i^s \sim N(\mu_\tau \cdot \widetilde{\sigma}(x_i;\beta^0), \widetilde{\sigma}(x_i;\beta^0))$ for $s = 1,\ldots,S$, and Ω is a positive definite weighting matrix. For the sake of brevity, these distributions $N(\mu_\tau \cdot \widetilde{\sigma}(x_i;\beta), \widetilde{\sigma}(x_i;\beta))$ will be referred to as $\widetilde{F}_{\widetilde{\varepsilon}(\beta)}$ within the binding function and its density will be denoted as $\widetilde{f}_{\widetilde{\varepsilon}(\beta)}$. The corresponding quantities for the response variable are $\widetilde{F}_{\widetilde{y}(\beta)}$ and $\widetilde{f}_{\widetilde{y}(\beta)}$.

To define a feasible indirect inference estimator, the simulated error distribution defined by $\widetilde{\sigma}(x_i;\beta)$ has to be estimated by $N(\mu_\tau \cdot \widehat{\sigma}_n(x_i;\beta^0), \widehat{\sigma}_n(x_i;\beta^0))$ using an estimate $\widehat{\sigma}_n(x_i;\beta)$. Denoting $\widehat{\theta}_n$ the linear QR estimate for the original data and $\widehat{F}_{y,n}(\cdot|x_i)$ an estimate of $F_y(\cdot|x_i)$, we define

$\widehat{\sigma}_n(x_i;\beta)$ as

$$\widehat{\sigma}_n(x_i;\beta) = \begin{cases} \frac{x_i^T\widehat{\theta}_n - x_i^T\beta}{\Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))} & \text{if } x_i^T\beta \le 0, \\ \frac{x_i^T\widehat{\theta}_n - x_i^T\beta}{\Phi_\tau^{-1}(\min\{\max\{\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i) - \widehat{F}_{y,n}(x_i^T\beta|x_i)+\tau),0\},1\})} & \text{if } x_i^T\beta > 0. \end{cases} \tag{16}$$

Since the denominators in (16) might take value 0, we again extend the definition (16) of $\widehat{\sigma}_n(x_i;\beta)$ in the such a way that the variance function $\widehat{\sigma}_n(x_i;\beta)$ is continuous is x_i. For a given β and any sequence $\{c_n\}_{n=1}^{\infty}$ such that $c_n = O(n^{-k_0})$, $k_0 > 0$, suppose that $|\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x) - \widehat{F}_{y,n}(x_i^T\beta|x)| < c_n$ for $x_i^T\beta > 0$ or $|\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i) - \tau| < c_n$ for $x_i^T\beta \le 0$; we refer to this event as the "zero-denominator" $ZD_{i,n}(\widehat{\theta}_n,\beta)$. If $ZD_{i,n}(\widehat{\theta}_n,\beta)$ occurs, then we use instead of (16) the linearly interpolated values

$$\widehat{\sigma}_n(x_i;\beta) = \frac{x_i^T\widehat{\theta}_n - x_m^T\widehat{\theta}_n}{x_M^T\widehat{\theta}_n - x_m^T\widehat{\theta}_n}\widehat{\sigma}_n(x_M;\beta) + \frac{x_M^T\widehat{\theta}_n - x_i^T\widehat{\theta}_n}{x_M^T\widehat{\theta}_n - x_m^T\widehat{\theta}_n}\widehat{\sigma}_n(x_m;\beta), \tag{17}$$

where $m = \arg\max_{j\le n}\{x_j^T\widehat{\theta}_n : x_j^T\widehat{\theta}_n < x_i^T\widehat{\theta}_n$ and $I(DZ_{j,n}(\widehat{\theta}_n,\beta)) = 0\}$ and $M = \arg\min_{j\le n}\{x_j^T\widehat{\theta}_n : x_j^T\widehat{\theta}_n > x_i^T\widehat{\theta}_n$ and $I(DZ_{j,n}(\widehat{\theta}_n,\beta)) = 0\}$; if $m = \emptyset$ or $M = \emptyset$ (e.g., if $\widehat{\theta}_n = \beta$), $\widehat{\sigma}_n(x_m;\beta) = 1$ or $\widehat{\sigma}_n(x_M;\beta) = 1$, respectively. Alternatively, one can also use a straightforward analog of (14) and define

$$\widehat{\sigma}_n(x_i;\beta) = \frac{\phi_\tau\{\Phi_\tau^{-1}[\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i)]\}}{\hat{f}_{y,n}(x_i^T\widehat{\theta}_n|x_i)}, \tag{18}$$

where $\hat{f}_{y,n}(\cdot|x_i)$ is an estimate of the conditional density function $f_y(\cdot|x_i)$. As this requires an additional nonparametric estimator, we rely in the theoretical part on definition (17) to minimize the number of required assumptions.

Having an estimate $\widehat{\sigma}_n(x_i;\beta)$ defined by (16)–(17) (or (16) and (18)), the feasible indirect inference (FII) estimator $\widehat{\beta}_n^{FII}$ can be defined as

$$\widehat{\beta}_n^{FII} = \arg\min_{\beta\in B}\left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}_n^s(\beta)\right]^T \Omega \left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}_n^s(\beta)\right], \tag{19}$$

where $\widehat{\theta}_n^s(\beta) = \arg\min_{\theta\in\Theta}\sum_{i=1}^{n}\rho_\tau(\widehat{y}_i^s(\beta) - x_i^T\theta)$ using $\widehat{y}_i^s(\beta) = \max\{0, x_i^T\beta + \widehat{\varepsilon}_i^s\}$ and $\widehat{\varepsilon}_i^s \sim N(\mu_\tau \cdot \widehat{\sigma}_n(x_i;\beta), \widehat{\sigma}_n(x_i;\beta))$.

Among final remarks on the proposed estimator $\widehat{\beta}_n^{FII}$, it does not perform a selection procedure as it is done in Khan and Powell (2001) and Chernozhukov and Hong (2002), that is, the proposed estimation method is applied to all observations in the sample. Furthermore, our estimation procedure corrects the downward bias of linear QR caused by the censoring of the dependent variable and can be thus considered as a bias-correction method. The bias-correction procedure is based, similarly to two-step estimators, on nonparametric estimates. Even though the bias-correction does not seem to be overly sensitive to the (lack of) precision of these nonparametric estimates, it could benefit from using some dimension reduction technique (e.g., Xia et al., 2002) to estimate $\sigma(x_i, \beta)$ on a lower dimensional space in models with a large numbers of explanatory variables, especially discrete ones.

4. Large sample properties

In this section, the asymptotic properties of the indirect-inference estimators for the censored regression model, $\widehat{\beta}_n^{III}$ and $\widehat{\beta}_n^{FII}$, are derived. As our main result, we prove that $\widehat{\beta}_n^{III}$ and $\widehat{\beta}_n^{FII}$ are asymptotically equivalent and asymptotically normally distributed. Let us first introduce conditions required for establishing the consistency and asymptotic normality of the III estimator.

A.1 The parameter spaces $\mathit{\Theta}$ and B are compact subsets of R^k and the true parameter values are $\theta^0 \in \Theta^\circ$ and $\beta^0 \in B^\circ$.

A.2 The parameter vector θ^0 uniquely solves $E\left[\varphi(y_i - x_i^T\theta) \cdot x_i\right] = 0$.

A.3 The random vectors $\{(x_i, y_i)\}_{i=1}^n$ are independent and identically distributed with finite second moments. The support of $x_i \in X$ is assumed to be compact. Moreover, the index $x_i^T\theta^0$ is continuously distributed, that is, there is at least one continuously distributed explanatory variable with $\theta_j^0 \neq 0$.

A.4 The τth conditional quantile of ε_i is zero. The error term ε_i has the conditional distribution $F_\varepsilon(t|x_i)$ with the conditional density function $f_\varepsilon(t|x_i)$, which is uniformly bounded both in t and x_i, positive on its support, and uniformly continuous with respect to t and x_i.

A.5 The following matrices are assumed to be are finite and positive definite:

- $J_{crq} = E\left[I(x_i^T\theta^0 > 0)I(x_i^T\beta^0 > 0)f_y(x_i^T\beta^0|x_i)x_ix_i^T\right]$,
- $J = E[I(x_i^T\theta^0 > 0)f_y(x_i^T\theta^0|x_i)x_ix_i^T]$,
- $\widetilde{J} = E\left[I(x_i^T\theta^0 > 0)\widetilde{f}_{\widetilde{y}(\beta^0)}(x_i^T\theta^0|x_i)x_ix_i^T\right]$,
- $\Sigma = E[\{\tau - I(y_i < x_i^T\theta^0)\}\{\tau - I(y_i < x_i^T\theta^0)\}x_ix_i^T]$,

- $\widetilde{\Sigma} = E[\{\tau - I(\widetilde{y}_i^s(\beta^0) < x_i^T\theta^0)\}\{\tau - I(\widetilde{y}_i^s(\beta^0) < x_i^T\theta^0)\}x_i x_i^T]$, and
- $K = E[(F_y(x_i^T\theta^0|x_i) - \tau)(F_y(x_i^T\theta^0|x_i) - \tau)x_i x_i^T]$.

A.6 Denoting $\widetilde{F}^0 = \widetilde{F}_{\widetilde{y}(\beta^0)}$, the link function $b(\widetilde{F}^0, \beta)$ is a one-to-one mapping. Moreover, $b(F, \beta)$ is assumed to be continuous in β and F (with respect to the supremum norm) at β^0 and $\widetilde{F}^0$. Finally, $b(\widetilde{F}_{\widetilde{y}(\beta)}, \beta)$ is continuously differentiable in $\beta \in U(\beta^0, \delta_b), \delta_b > 0$, and $D = \partial b(\widetilde{F}_{\widetilde{y}(\beta^0)}, \beta^0)/\partial\beta^T$ has a full column rank.

A.7 $P(\Delta(x_i, \theta^0, \beta^0) = 0) = 0$ and $P(x_i^T\theta^0 = c) = 0$ for any $c \in R$.

Let us provide a few remarks regarding the necessity of these assumptions. Assumptions A.1, A.2, and A.3 are essential for establishing the consistency and asymptotic normality of the QR estimates $\widehat{\theta}_n$ as argued in Angrist et al. (2006) as well as the consistency and asymptotic normality of $\widehat{\beta}_n^{III}$. As shown in Angrist et al. (2006), compactness of the support of X, which is also required by Khan and Powell (2001), can be relaxed to the existence of finite $(2+\delta)$th moment of x_i. In our proofs of the asymptotic properties of $\widehat{\beta}_n^{III}$, relaxing the compactness of X would additionally require $\max_{i\leq n}\|x_i\| = o_p(n^\alpha)$ for some $0 < \alpha < 1/2$. This is however an assumption closely related to the existence of finite $(2+\delta)$th moments (cf. Čížek, 2006, Proposition 2.1). Moreover, we assume the existence of one continuous explanatory variable. This assumption is not strictly necessary for the validity of the asymptotic results, but it is needed for constructing a practically applicable formula for the asymptotic variance of the misspecified QR and FII estimators; with only discrete variables, an alternative approaches such as bootstrap would have to be used to compute the asymptotic variance.

Next, Assumption A.4 is the standard assumption in quantile regression models (e.g., Powell, 1986a), although the density function $f_\varepsilon(t|x_i)$ is usually assumed to be positive only in a neighborhood of 0. Given the misspecification of the linear QR, it is convenient to assume non-zero density everywhere as $f_\varepsilon(t|x_i)$ is evaluated for any $t = x_i^T\theta^0$. Concerning Assumption A.5, it contains usual full-rank conditions used in censored and quantile regression and is necessary for the identification of parameter vectors, see for example Khan and Powell (2001). (As the QR slope estimates are typically biased towards zero under censoring from below, $x_i^T\beta^0 > 0$ usually implies $x_i^T\theta^0 > 0$ and assumptions $J > 0$ and $\widetilde{J} > 0$ are thus weaker than $J_{crq} > 0$.) Further, Assumption A.6 is the standard assumption necessary for defining the indirect inference estimator: the population QR estimates $\widetilde{\theta}(\beta)$ and $\widetilde{\theta}'(\beta')$ for data simulated from the censored regression model with parameters β and β' should differ if $\beta \neq \beta'$. Note though that we require the link function to be one-to-one only at the distribution $\widetilde{F}^0 = \widetilde{F}_{\widetilde{y}(\beta^0)}$ corresponding to the true parameter values β^0; alternatively to A.6, $b(F_y, \beta)$ can be assumed to be one-to-one instead. Finally, the first part of Assumption A.7 is imposed to simplify the proof of the consistency and asymptotic normality of the proposed estimator: it rules out the data without any censoring. The results remain valid even if there is no censoring, although some proofs would slightly differ. The second part of Assumption A.7 just formalizes the continuous-regressor Assumption A.3.

These assumptions are sufficient to derive the asymptotic distribution of the infeasible estimator. For the sake of simplicity of some proofs, we will additionally assume that the conditional error distribution $F_\varepsilon(\cdot|x_i)$ has an infinite support (see Appendix A for details), but the stated results are valid in the general case as well.

Theorem 2. *Let quantile $\tau \in (0,1)$, Ω be a non-singular $k \times k$ matrix, and $S \in N$ be a fixed number of simulated samples. Under Assumptions A.1–A.7, $\widehat{\beta}_n^{III}$ is a consistent estimator of β^0 and it is asymptotically normal:*

$$\sqrt{n}(\widehat{\beta}_n^{III} - \beta^0) \to N\left(0, D^{-1}V(S)(D^T)^{-1}\right) \tag{20}$$

as $n \to +\infty$, where $V(S) = J^{-1}\Sigma J^{-1} + \frac{1}{S}\widetilde{J}^{-1}\widetilde{\Sigma}\widetilde{J}^{-1} + (1-\frac{1}{S})\widetilde{J}^{-1}\widetilde{K}\widetilde{J}^{-1} - 2J^{-1}\widetilde{K}\widetilde{J}^{-1}$.

Proof: See Appendix C. □

The asymptotic variance matrix of $\widehat{\beta}_n^{III}$ derived in Theorem 2 consists of several parts. First, the matrices Σ and $\widetilde{\Sigma}$ are the variances of the QR first-order conditions in the real and simulated data, respectively. Next, J and $\widetilde{J}$ are the corresponding Jacobian matrices defined in Assumption A.5. Finally, matrix K characterizes the unconditional covariance between the real and simulated data.

The next theorem shows that the feasible estimator $\widehat{\beta}_n^{FII}$ is asymptotically equivalent to the infeasible one $\widehat{\beta}_n^{III}$ provided that one extra assumption holds: the conditional distribution function and its nonparametric estimates, which are used in (16) to define $\widehat{\sigma}_n(x_i;\beta)$, have to be smooth functions of data. Additionally, the nonparametric estimate $\widehat{F}_{y,n}(z_i|x_i)$ has to be consistent and converge at a faster rate than the sequence $c_n = O(n^{-k_0})$, $k_0 > 0$, used in the definition of $\widehat{\sigma}_n(x_i;\beta)$. This is however not a constraint as k_0 is arbitrary.

A.8 Estimator $\widehat{F}_{y,n}(t|x_i)$ is a Lipschitz function in $t \in R^+$ uniformly in x_i, and for any compact sets $C_x \subset R^k$ and $C_t \subset R^+$, $\sup_{x \in C_x} \sup_{t \in C_t} |\widehat{F}_{y,n}(t|x_i) - F_y(t|x_i)| = O_p(n^{-k_1})$ for some $k_1 > k_0 > 0$. Moreover, $\widehat{F}_{y,n}(t|x_i) = 0$ for any $t < 0$ and $x_i \in X$.

A.9 The conditional distribution functions $F_y(z|x_i = x)$ are piecewise Lipschitz functions in x for any $z \in R$.

Assumption A.8 is satisfied for many commonly used estimators of conditional distribution functions. Assumption A.9 on the conditional distribution function then states explicitly a minimum requirement that facilitates a consistent estimation and hence validity of Assumption A.8, although stronger assumptions on the smoothness of $F_y(z|x_i)$ are usually used (cf. Li and Racine, 2004).

Theorem 3. *Let the assumptions of Theorem 2 be satisfied. If Assumptions A.8–A.9 also hold,* $\sqrt{n}(\widehat{\beta}_n^{FII} - \widehat{\beta}_n^{III}) \to 0$ *in probability as* $n \to +\infty$.

Proof: See Appendix C. □

Theorem 3 shows that the feasible and infeasible II estimates, $\widehat{\beta}_n^{FII}$ and $\widehat{\beta}_n^{III}$, are asymptotically equivalent, and consequently, the asymptotic variance-covariance matrix of $\widehat{\beta}_n^{FII}$ is given in by (20). Note that all elements of matrix $V(S)$ can be readily estimated in practice (after replacing θ^0 by $\widehat{\theta}_n$) as most nonparametric estimates of the conditional distribution functions and densities are computed already during the estimation: they define conditional variances $\widehat{\sigma}_n(x_i, \beta)$ (a possible exception is $f_y(\cdot|x_i)$ depending on the employed definition of $\widehat{\sigma}_n(x_i, \beta)$). Note that the simulated distribution $\widetilde{F}_{\widetilde{y}(\beta^0)}$ and density $\widetilde{f}_{\widetilde{y}(\beta^0)}$ functions do not have to be estimated after obtaining $\widehat{\beta}_n^{FII}$ either as they are normal and defined by μ_τ and $\widetilde{\sigma}(x_i, \beta^0)$, which is consistently estimated by $\widehat{\sigma}_n(x_i, \widehat{\beta}_n^{FII})$ (see Lemma 8). Additionally, the variance matrix of the proposed estimator depends on the derivative D of the link function. As this derivative is defined in terms of the simulated model, the link function and its derivative can be easily estimated by simulating a sufficient number of samples from the model (1)–(2) with parameter $\widehat{\beta}_n^{FII}$ and by computing derivatives numerically as discussed in Gourieroux and Monfort (1993).

5. Monte Carlo simulations

Although we characterized the asymptotic properties of the proposed bias-corrected QR estimator, it is primarily aimed to improve the finite-sample performance of existing estimators (e.g., the heavy-tailed distribution of CLAD in small samples). To analyze the benefits of the bias-correction performed by means of the indirect inference, this method is now compared with many existing estimators by means of Monte Carlo simulations. The simulation setting is described in Section 5.1 and the results are discussed in Section 5.2.

5.1. Simulation design

The data-generating process is very similar to the one considered by Khan and Powell (2001). We consider

$$y_i = \max\{\alpha + \beta x_i + \varepsilon_i, 0\}$$

with the slope parameter $\beta \in R$ and a univariate regressor x_i; as the results are qualitatively rather similar across different data distributions, we report results for x_i being uniformly distributed on $\langle -\sqrt{3}, \sqrt{3} \rangle$. Similarly to Khan and Powell (2001), $\beta = 1$ and α is chosen in each sample so that the censoring level stays always equal to 50%. Further, we focus on the median regression case $\tau = 0.5$. The error term ε_i can thus follow various error distributions with median equal to zero, such as the normal $N(0, \sigma_x)$, Student t_d, and double exponential $DExp(\lambda)$ ones.

For this data-generating process, we consider the following estimators: (i) the standard Tobit maximum likelihood estimator (MLE) constructed for normal homoscedastic errors; (ii) the CLAD estimator computed by the exhaustive search for its global minimum; (iii) the two-step LAD of Khan and Powell (2001) is computed based on their three initial estimators – the maximum score estimator (2S-MSC), the Nadaraya-Watson estimator of the propensity score (2S-NW), and the conditional quantile estimator (2S-LQR); (iv) the 'infeasible' LAD (IFLAD) computed by quantile regression applied only to data points with $\alpha + \beta x_i \geq 0$; (v) the three-step estimator of Chernozhukov and Hong (2002) based on the initial logit estimator (3S-LOG); (vi) the proposed QR estimator with bias corrected by indirect inference (FII); and (vii) the corresponding infeasible indirect inference (III) estimator, which does not estimate the conditional error distribution, but 'knows' the true one.

The QR estimates were in all cases computed by the Barrodale and Roberts algorithm as implemented in the R package "quantreg." The same package was also used for computing CLAD by means of Fitzenberger (1997b)'s algorithm. For the indirect-inference based methods, we use the Nelder-Mead simplex method with multiple starting points as an optimization algorithm. Further, many of the considered methods depend on some initial nonparametric estimators of the conditional mean, conditional quantile, and conditional distribution and density functions. The nonparametric estimators considered here are those by Racine and Li (2004) for the conditional mean and distribution function estimation, by Li and Racine (2008) for the conditional quantile estimation, and by Hall et al. (2004) for the conditional density estimation; we use their implementation in the R package "npreg," which also includes the bandwidth choice by the least-squares cross-validation. The estimation and the bandwidth choice were done in all cases by the k-nearest-neighbors estimation with the uniform kernel. Finally, the bias-corrected QR estimator using the estimated values of the conditional distribution function can sometimes exhibit multiple minima in small samples (in such cases, there are usually two minima found irrespective of the number of starting points used in the Nelder-Mead optimization algorithm): we then choose the one, which commands the smallest sum of absolute residuals for observations with positive values of the (nonparametrically estimated) conditional median $\mathrm{med}(y_i|x_i)$. The number of simulated samples is $S = 400$ in all cases.

The results for all methods are obtained using 1000 simulations for sample sizes $n = 50, 100$, and 200 and are summarized using the bias and root mean squared error (RMSE) of the slope estimates.

5.2. Simulation results

The first set of results is obtained for three different error distributions, which are homoscedastic in all cases: $N(0, 1)$, t_5, and $DExp(1)$, see Table 1. The MLE estimator serves as a parametric benchmark. First, CLAD exhibit large biases and RMSEs in small samples with $n = 100$ and especially with $n = 50$ observations; this is due to the heavy right-tail of the CLAD distribution. Next, the existing two-step estimators exhibit relatively large RMSEs, which are however always smaller than those of CLAD, and negative biases, which vary with the choice of the initial estimator. In comparison, the three-step estimator 3S-LOG has usually slightly larger RMSE than 2S-NW or 2S-LQR, but possesses rather small finite-sample bias compared to all other semiparametric methods. All these existing methods are asymptotically equivalent to the infeasible LAD, which is naturally more precise (but infeasible) in finite samples.

Looking at the two infeasible estimators, III exhibits always smaller RMSE than IFLAD, although the difference decreases with an increasing sample size. It is also interesting to note that IFLAD exhibits systematically a larger negative bias, whereas III leads to a smaller, but positive bias (or almost zero bias for $n = 200$). This is reflected by the performance of the proposed FII estimator, which exhibits much smaller bias and RMSE for $n = 50$ and which is preferable to any of the existing methods in all cases with the exception of $n = 200$ and the double-exponential errors, where 2S-LQR has a smaller RMSE, but a larger bias. One can also notice that, even though FII exhibits generally a smaller bias than the methods of Khan and Powell (2001), the bias of FII is negative in contrast to the bias of III.

The second set of results is obtained for normally distributed errors with different types of heteroscedasticity, $N(0, \sigma_x)$, where $\sigma_x = 1$ in the case of homoscedasticity, $\sigma_x = ce^{0.75x}$ in the case of "positive" heteroscedasticity, and $\sigma_x = ce^{-0.75x}$ in the case of "negative" heteroscedasticity (c is always chosen so that the unconditional variance equals 1). Similarly to the homoscedastic case, CLAD exhibits extreme bias and RMSE in small samples, but provides good estimates if $n = 200$ (this is however a large sample size considering that only one explanatory variable is present). The existing two- and three-step estimators provide better estimates than CLAD in almost all cases, but the biggest gain is observed in the case of positive heteroscedasticity.

Comparing now the existing methods and the proposed bias-corrected QR based on II, the infeasible estimator IFLAD is now inferior to III only in the case of homoscedasticity and positive heteroscedasticity, while III performs worse than IFLAD if negative heteroscedasticity is used. This is a consequence of III employing all observations and IFLAD using only those with $\alpha + \beta x_i > 0$ ($\beta = 1$): in the case of negative heteroscedasticity, IFLAD thus uses only observations with the smallest conditional variance, and consequently, is more precise than III using all observations (including those with large conditional variance). This is reflected by the results of the feasible FII estimator, which performs similarly across data designs: it outperforms all existing methods for the homoscedastic data and data with positively heteroscedastic errors, where the difference is largest at small sample sizes. In the case of negative heteroscedasticity, FII has RMSEs comparable to those of 2S-NW, for instance, and worse than 2S-LQR for $n \geq 100$ and than 3S-LOG at any sample size (note that 3S-LOG together with CLAD seem to be the most sensitive methods to the changes in the structure of conditional variances).

Altogether, all semiparametric alternatives to CLAD perform better than CLAD, although the differences are likely to be small for very large samples. The proposed FII estimator performs equally well in large samples and is almost always preferable to all existing semiparametric methods in small samples.

Table 1: The bias and root mean squared error of all estimators in a model with the standard normal, Student, and double exponential errors.

Sample size	Estimator	$\varepsilon \sim N(0,1)$ Bias	$\varepsilon \sim N(0,1)$ RMSE	$\varepsilon \sim t_5$ Bias	$\varepsilon \sim t_5$ RMSE	$\varepsilon \sim DExp(1)$ Bias	$\varepsilon \sim DExp(1)$ RMSE
$n = 50$	MLE	-0.006	0.205	0.047	0.267	0.079	0.300
	CLAD	0.527	2.971	0.937	5.398	3.464	50.59
	IFLAD	*-0.127*	*0.473*	*-0.128*	*0.497*	*-0.110*	*0.487*
	2S-MSC	-0.259	0.568	-0.281	0.611	-0.254	0.592
	2S-NW	-0.167	0.487	-0.182	0.535	-0.149	0.500
	2S-LQR	-0.150	0.469	-0.162	0.501	-0.145	0.476
	3S-LOG	-0.037	0.510	-0.036	0.527	-0.024	0.526
	III	*0.041*	*0.311*	*0.051*	*0.366*	*0.060*	*0.376*
	FII	-0.057	0.371	-0.072	0.374	-0.077	0.357
$n = 100$	MLE	-0.004	0.142	0.052	0.199	0.079	0.216
	CLAD	0.183	0.584	0.227	1.086	0.169	0.652
	IFLAD	*-0.074*	*0.295*	*-0.086*	*0.322*	*-0.071*	*0.292*
	2S-MSC	-0.167	0.367	-0.206	0.424	-0.167	0.384
	2S-NW	-0.105	0.332	-0.115	0.369	-0.084	0.332
	2S-LQR	-0.094	0.322	-0.096	0.342	-0.077	0.319
	3S-LOG	-0.016	0.373	-0.035	0.381	-0.011	0.360
	III	*0.018*	*0.202*	*0.019*	*0.226*	*0.042*	*0.250*
	FII	-0.051	0.277	-0.033	0.302	-0.036	0.285
$n = 200$	MLE	0.003	0.101	0.063	0.174	0.079	0.159
	CLAD	0.068	0.301	0.080	0.304	0.060	0.269
	IFLAD	*-0.059*	*0.221*	*-0.067*	*0.236*	*-0.042*	*0.178*
	2S-MSC	-0.128	0.271	-0.125	0.271	-0.115	0.236
	2S-NW	-0.076	0.242	-0.066	0.248	-0.060	0.209
	2S-LQR	-0.058	0.228	-0.052	0.232	-0.049	0.193
	3S-LOG	-0.020	0.265	-0.008	0.281	-0.012	0.237
	III	*-0.008*	*0.137*	*0.011*	*0.141*	*0.009*	*0.162*
	FII	-0.031	0.220	-0.021	0.213	-0.023	0.206

Table 2: The bias and root mean squared error of all estimators in a model with the normal errors under homoscedasticity and positive and negative heteroscedasticity.

Sample size	Estimator	$\sigma_x = 1$		$\sigma_x \sim e^{0.75x}$		$\sigma_x \sim e^{-0.75x}$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
$n = 50$	MLE	-0.006	0.205	0.595	0.703	-0.448	0.479
	CLAD	0.527	2.971	21.54	482.1	-0.023	3.032
	IFLAD	*-0.127*	*0.473*	*-0.072*	*0.510*	*-0.036*	*0.227*
	2S-MSC	-0.259	0.568	-0.133	0.623	-0.169	0.310
	2S-NW	-0.167	0.487	-0.050	0.544	-0.162	0.340
	2S-LQR	-0.150	0.469	-0.062	0.541	-0.164	0.308
	3S-LOG	-0.037	0.510	-0.134	0.809	-0.069	0.235
	III	*0.041*	*0.311*	*0.029*	*0.490*	*0.063*	*0.308*
	FII	-0.057	0.371	-0.015	0.436	-0.104	0.279
$n = 100$	MLE	-0.004	0.142	0.614	0.671	-0.448	0.468
	CLAD	0.183	0.584	2.145	53.96	0.026	0.190
	IFLAD	*-0.074*	*0.295*	*-0.045*	*0.351*	*-0.042*	*0.161*
	2S-MSC	-0.167	0.367	-0.099	0.431	-0.125	0.213
	2S-NW	-0.105	0.332	-0.044	0.399	-0.105	0.227
	2S-LQR	-0.094	0.322	-0.040	0.401	-0.106	0.197
	3S-LOG	-0.016	0.373	0.014	0.569	-0.073	0.164
	III	*0.018*	*0.202*	*0.009*	*0.284*	*0.030*	*0.184*
	FII	-0.051	0.277	-0.005	0.328	-0.065	0.214
$n = 200$	MLE	0.003	0.101	0.619	0.646	0.435	0.443
	CLAD	0.068	0.301	0.102	0.401	0.006	0.134
	IFLAD	*-0.059*	*0.221*	*-0.024*	*0.236*	*-0.033*	*0.110*
	2S-MSC	-0.128	0.271	-0.069	0.281	-0.093	0.159
	2S-NW	-0.076	0.242	-0.020	0.269	-0.080	0.164
	2S-LQR	-0.058	0.228	-0.015	0.270	-0.081	0.137
	3S-LOG	-0.020	0.265	-0.013	0.418	-0.075	0.127
	III	*-0.008*	*0.137*	*0.012*	*0.190*	*0.019*	*0.121*
	FII	-0.031	0.220	0.010	0.250	-0.046	0.163

6. Conclusion

We proposed a new estimation method for the censored regression models that – contrary to existing methods – relies on the linear regression QR estimates for the whole sample and that applies a bias-correction technique to obtain consistent estimates. For the bias correction, the indirect inference technique is applied and extended so that it allows sampling from a nonparametrically estimated distribution function. The consistency and asymptotic distribution of the proposed estimator were found and shown to be first-order independent of the initial nonparametric estimates of the auxiliary error distribution. Finally, one of important benefits of this estimation approach is its small-sample performance as was demonstrated by means of Monte Carlo simulations.

Appendix A. Proof of Theorem 1

Proof of Theorem 1: The proof is almost identical to the proof of Theorem 2 in Angrist et al. (2006). We need to prove that the solution of

$$\theta^0 = \arg\min_{\theta \in \Theta} E\left[\rho_\tau(y_i - x_i^T \theta)\right] \tag{21}$$

is equal to the solution of

$$\bar{\theta} = \arg\min_{\theta \in \Theta} E\left[w(x_i, \beta^0, \bar{\theta}) \cdot \Delta^2(x_i, \beta^0, \theta)\right]. \tag{22}$$

Since the objective function in (22) is convex, any fixed point $\theta = \bar{\theta}$ is a solution of the corresponding first-order condition:

$$\mathcal{F}(\theta) = 2 \cdot E\left[w(x_i, \beta^0, \theta) \cdot \Delta(x_i, \beta^0, \theta) \cdot x_i\right] = 0. \tag{23}$$

On the other hand, the first order condition for (21) is given by (cf. the proof of Theorem 2 in Angrist et al., 2006)

$$\mathcal{D}(\theta) = E\left[I\left\{\left(u_i \leq \Delta(x_i, \beta^0, \theta)\right) - \tau\right\} \cdot x_i\right] = 0. \tag{24}$$

By the law of iterated expectations, $D(\theta)$ can be written as

$$\mathcal{D}(\theta) = E\left[\{F_{u_i}(\Delta(x_i, \beta^0, \theta)|x_i) - \tau\} \cdot x_i\right] = 0. \tag{25}$$

Since $F_{u_i}(\Delta(x_i, \beta^0, \theta)|x_i) = \tau$ if $\Delta\left(x_i, \beta^0, \theta\right) = 0$, it follows from the definition of $w(x_i, \beta^0, \theta)$ that

$F_{u_i}(\Delta(x_i,\beta^0,\theta)|x_i) - \tau = w(x_i,\beta^0,\theta)\cdot\Delta(x_i,\beta^0,\theta)$ for any value of $\Delta\left(x_i,\beta^0,\theta\right)$, and consequently,

$$\mathcal{D}(\theta) = E\left[\{F_{u_i}(\Delta(x_i,\beta^0,\theta)|x_i) - \tau\}\cdot x_i\right] = 2\cdot E\left[w(x_i,\beta^0,\theta)\cdot\Delta(x_i,\beta^0,\theta)\cdot x_i\right] = \mathcal{F}(\theta).$$

Because θ^0 is the unique solution of (21), it also uniquely solves (22) since the objective function in (22) is convex in θ. Therefore, $\theta = \theta^0 = \bar{\theta}$ solves both (21) and (22). $\square$

Appendix B. Auxiliary lemmas

First, we introduce necessary notation. The norms $\|\cdot\|$ and $\|\cdot\|_\infty$ will refer to the Euclidean norm on R^d and to the supremum norm in functional spaces, respectively. The δ-neighborhood of a vector $t\in R^d$ is denoted $U(t,\delta) = \{t'\in R^d : \|t'-t\| < \delta\}$. The probability distribution and density functions of $N(\mu_\tau,1)$ are denoted Φ_τ and ϕ_τ, respectively. Additionally, recall that $\rho_\tau(z) = \{\tau - I\left(z\le 0\right)\}\cdot z$ and its derivative is denoted $\varphi_\tau(z) = \tau - I(z\le 0)$.

Next, for $w_i = x_i$ or $w_i = (y_i,x_i)$, let $E_n[f(w_i)]$ denote $n^{-1}\sum_{i=1}^{n} f(w_i)$ and let $G_n[f(w_i)]$ denote $n^{-1/2}\sum_{i=1}^{n}\{f(w_i) - E[f(w_i)]\}$. If we need to indicate a particular data distribution P of w_i, $G_{n,P}[f(w_i)] = n^{-1/2}\sum_{i=1}^{n}\{f(w_i) - E_P[f(w_i)]\}$ is used, assuming that $w_i\sim P$. For easier reading, we also use a simplified notation for the simulated distributions $\widetilde{F}(\beta) = \widetilde{F}_{\widetilde{y}(\beta)}$ and $\widehat{F}(\beta) = \widehat{F}_{\widehat{y}(\beta)}$.

For the sake of simplicity of some proofs, we will additionally assume that the conditional error distribution $F_\varepsilon(\cdot|x_i)$ has (uniformly) an infinite support in order to guarantee that $\sup_{x\in X} F_\varepsilon(K|x) < 1$ for any $K<\infty$, and by Assumption A.3, that $\sup_{x\in X}\sup_{\theta\in\Theta} F_y(x^T\theta|x) < K_F < 1$. Consequently, the conditional variance $\widetilde{\sigma}(x_i;\beta^0)$ defined in (13)–(14) is everywhere positive at the true β^0, and given the compactness of B, Θ, X, and A.7, $\widetilde{\sigma}(x_i;\beta^0) > C_\sigma > 0$ for all $x_i\in X$. If the limit expression (14) and Assumption A.4 are taken into account, one can observe that the variance function is also bounded from above: $\widetilde{\sigma}(x_i;\beta) < K_\sigma$ for any $\beta\in B$ and all $x_i\in X$.

Lemma 4. *Under Assumptions A.1–A.5 and A.7, there is some $\delta > 0$ such that $-\partial E[\varphi_\tau(y_i - x_i^T\theta)x_i]/\partial\theta = \partial E[\{F_y(x_i^T\theta|x_i) - \tau\}x_i]/\partial\theta = E[I(x_i^\top\theta > 0)f_y(x_i^T\theta|x_i)x_ix_i^T]$ for any $\theta\in U(\theta^0,\delta)$. Moreover, $E[I(x_i^\top\theta^0 > 0)f_y(x_i^T\theta^0|x_i)x_ix_i^T]$ is a positive definite matrix.*

Proof: By Assumption A.7, $P(x_i^T\theta^0 = 0) = 0$ and there exists some $\delta > 0$ such that $P(x_i^T\theta = 0) = 0$ for any $\theta\in U(\theta^0,\delta)$. Note that $-E[\varphi_\tau(y_i - x_i^T\theta)x_i] = E[\{I(y_i - x_i^T\theta) - \tau\}x_i] = E[(E\{I(y_i - x_i^T\theta)|x_i\} - \tau)x_i] = E[\{F_y(x_i^T\theta|x_i) - \tau\}x_i]$ by the law of iterated expectations.

Fixing a particular $\theta' \in U(\theta^0, \delta)$ and rewriting

$$
\begin{aligned}
E[\{F_y(x_i^T\theta|x_i) - \tau\}x_i] &= E[I(x_i^T\theta' < 0) \cdot \{F_y(x_i^T\theta|x_i) - \tau\}x_i] \qquad (26)\\
&+ E[I(x_i^T\theta' = 0) \cdot \{F_y(x_i^T\theta|x_i) - \tau\}x_i]\\
&+ E[I(x_i^T\theta' > 0) \cdot \{F_y(x_i^T\theta|x_i) - \tau\}x_i] \qquad (27)
\end{aligned}
$$

shows that we only have to differentiate terms (26) and (27) as the remaining term equals zero (see Assumption A.3).

First, consider the term (27):

$$
\begin{aligned}
\frac{\partial}{\partial\theta} E[I(x_i^T\theta' > 0) \cdot \{F_y(x_i^T\theta|x_i) - \tau\}x_i]\Big|_{\theta=\theta'} &= \frac{\partial}{\partial\theta} E_x(E[I(x_i^T\theta' > 0) \cdot \{F_y(x_i^T\theta|x_i) - \tau\}x_i|x_i])\Big|_{\theta=\theta'}\\
&= \frac{\partial}{\partial\theta} E_x(I(x_i^T\theta' > 0) \cdot E[\{F_y(x_i^T\theta|x_i) - \tau\}x_i|x_i])\Big|_{\theta=\theta'},
\end{aligned}
$$

which follows from the law of iterated expectation. If $I(x_i^T\theta' > 0) = 1$, then – conditionally on x_i – there is a positive constant c such that $x_i^T\theta' > c > 0$. An open neighborhood $U(\theta', \delta'), \delta > \delta' > 0$, can thus be found such that $c/2 < x_i^T\theta < 3c/2$ for all $\theta \in U(\theta', \delta')$. If x_i satisfies $I(x_i^T\theta' > c > 0) = 1$, $\partial E[\{F_y(x_i^T\theta'|x_i) - \tau\}x_i|x_i]/\partial\theta = E[f_y(x_i^T\theta'|x_i)x_i x_i^T|x_i]$, where the existence of the density f_y, and its continuity and boundedness follows from $y_i = \max\{0, x_i^T\theta^0 + \varepsilon_i\}$, Assumption A.4, and the fact that $x_i^T\theta > c/2 > 0$ for $\theta \in U(\theta', \delta')$. Noting that $E[f_y(x_i^T\theta|x_i)x_i x_i^T|x_i]$ is continuous in θ and uniformly bounded by Assumptions A.3 and A.4, it follows that

$$
\begin{aligned}
\frac{\partial}{\partial\theta} E[I(x_i^T\theta' > 0) \cdot \{F_y(x_i^T\theta|x_i) - \tau\}x_i]\Big|_{\theta=\theta'} &= E_x\left(I(x_i^T\theta' > 0) \cdot \frac{\partial}{\partial\theta} E[\{F_y(x_i^T\theta|x_i) - \tau\}x_i|x_i]\Big|_{\theta=\theta'}\right)\\
&= E_x(I(x_i^T\theta' > 0) \cdot E[f_y(x_i^T\theta'|x_i)x_i x_i^T|x_i])\\
&= E[I(x_i^T\theta' > 0) \cdot f_y(x_i^T\theta'|x_i)x_i x_i^T].
\end{aligned}
$$

By the same argument, (26) equals after differentiation to

$$
\begin{aligned}
\frac{\partial}{\partial\theta} E[I(x_i^T\theta' < 0) \cdot \{F_y(x_i^T\theta|x_i) - \tau\}x_i]\Big|_{\theta=\theta'} &= E_x\left(I(x_i^T\theta' < 0) \cdot \frac{\partial}{\partial\theta} E[\{F_y(x_i^T\theta|x_i) - \tau\}x_i|x_i]\Big|_{\theta=\theta'}\right)\\
&= 0
\end{aligned}
$$

as $F_y(t|x_i) = 0$ for any $t < 0$. The last claim of the theorem now follows from Assumption A.5. $\square$

Theorem 5. *Let $J = E[I(x_i^\top\theta > 0)f_y(x_i^T\theta|x_i)x_i x_i^T]$. Under Assumptions A.1–A.5 and A.7, it holds* *that*

1. $Q_n(\theta) = E_n[\rho_\tau(y_i - x_i^T\theta) - \rho_\tau(y_i - x_i^T\theta^0)] \to Q_\infty(\theta) = E[\rho_\tau(y_i - x_i^T\theta) - \rho_\tau(y_i - x_i^T\theta^0)]$ as $n \to \infty$ for any $\theta \in \Theta$;
2. $\widehat{\theta}_n$ *is a consistent estimator of* θ^0, $\widehat{\theta}_n \xrightarrow{P} \theta^0$ *as* $n \to \infty$*;*
3. *for any sequence* $\theta_n \xrightarrow{P} \theta^0$, $n^{1/2}(\theta_n - \theta^0) = -J^{-1}G_n\{\varphi_\tau(y_i - x_i^T\theta^0)x_i\} + o_p(1)$ *converges to a Gaussian process with covariance function* $\Sigma = E[(\tau - I(y_i < x_i^T\theta^0))(\tau - I(y_i < x_i^T\theta^0))x_ix_i]$.

Additionally, suppose that, for sample size $n \in N$*, data are independently and identically distributed according to probability distributions* P_n*, which satisfy Assumptions A.1–A.5 and A.7 uniformly in* n*. Denoting*

$$r(y, x, \theta) = \begin{cases} [\rho_\tau(y - x^T\theta) - \rho_\tau(y - x^T\theta^0) - (\theta - \theta^0)^T(\tau - I\{y \le x^T\theta^0\}) \cdot x]/\|\theta - \theta^0\| & \text{if } \theta \ne \theta^0, \\ 0 & \text{if } \theta = \theta^0, \end{cases}$$

let us assume that

$$\sup_{\theta', \theta'' \in U(\theta^0, \delta)} |E_{P_n}\{r(y_i, x_i, \theta') - r(y_i, x_i, \theta'')\}^2 - E_{P_0}\{r(y_i, x_i, \theta') - r(y_i, x_i, \theta'')\}^2| \to 0 \tag{28}$$

as $n \to \infty$ *for some* $\delta > 0$ *and some distribution* P_0*, which satisfies assumptions A.1–A.5 and A.7. Then for any sequence* $\theta_n \xrightarrow{P} \theta^0$, $n^{1/2}(\theta_n - \theta^0) = -J^{-1}G_{n,P_n}\{\varphi_\tau(y_i - x_i^T\theta^0)x_i\} + o_p(1)$ *as* $n \to \infty$.

Proof: To prove the consistency of $\widehat{\theta}_n$, we follow the same steps as in Angrist et al. (2006, Theorem 3). By definition (3), $\widehat{\theta}_n$ minimizes $Q_n(\theta) = E_n[\rho_\tau(y_i - x_i^T\theta) - \rho_\tau(y_i - x_i^T\theta^0)]$ with the corresponding limit denoted $Q_\infty(\theta) = E[\rho_\tau(y_i - x_i^T\theta) - \rho_\tau(y_i - x_i^T\theta^0)]$. As $\rho_\tau(t) = \{\tau - I(t < 0)\}t$, we obtain for any $\theta \in \Theta$

$$|Q_\infty(\theta)| \le 2 \cdot E\left|x_i^T(\theta - \theta^0)\right| \le 2 \cdot E\|x_i\| \cdot \|\theta - \theta^0\| < \infty. \tag{29}$$

By Khinchine's law of large numbers (Davidson, 1994, Theorem 23.5), $Q_n(\theta)$ converges pointwise to $Q_\infty(\theta)$ in probability as $n \to \infty$ for any $\theta \in \Theta$. Additionally, $Q_n(\theta)$ is stochastically equicontinuous because $|Q_n(\theta^{'}) - Q_n(\theta^{''})| \le C \cdot \|\theta^{'} - \theta^{''}\|$, where $C = 2 \cdot E\|x_i\| < \infty$ and $\theta^{'}, \theta^{''} \in \Theta$. Therefore, the uniform convergence of $Q_n(\theta)$ to $Q_\infty(\theta)$ in $\theta \in \Theta$ is established by Davidson (1994, Theorem 21.9), for instance. Since the θ^0 is a unique minimizer of $Q_\infty(\theta)$ by Assumption A.2, the consistency of $\widehat{\theta}_n$ follows from the standard consistency theorem for (convex) minimization problems (e.g., Newey and McFadden, 1994, Theorem 2.1 or Theorem 2.7).

To derive the asymptotic linearity of $\widehat{\theta}_n$ and its asymptotic distribution, we rely on Van der Vaart and Wellner (1996, Lemma 3.2.19), who derive the stochastic expansion of $n^{1/2}(\widehat{\theta}_n - \theta^0)$. Given Assumptions A.1 and A.3, Lemma 3.2.19 further requires that

1. The objective function $E[\rho_\tau(y_i - x_i^T\theta)]$ is twice differentiable at θ^0 with a non-singular second

derivative matrix J;

2. The criterion function $\rho_\tau(y_i - x_i^T\theta)$ is stochastically differentiable with the derivative equal to $\varphi_\tau(y_i - x_i^T\theta)x_i$ in the sense that

$$E[\rho_\tau(y_i - x_i^T\theta) - \rho_\tau(y_i - x_i^T\theta^0) - (\theta - \theta^0)^T\varphi_\tau(y_i - x_i^T\theta) \cdot x_i]^2 = o(\|\theta - \theta^0\|^2) \qquad (30)$$

3. The class of functions $\mathcal{G} = \{r(y, x, \theta) : \|\theta - \theta^0\| < \delta\}$ is Donsker for some $\delta > 0$, where

$$r(y, x, \theta) = \begin{cases} \frac{\rho_\tau(y - x^T\theta) - \rho_\tau(y - x^T\theta^0) - (\theta - \theta^0)^T(\tau - I\{y \le x^T\theta^0\}) \cdot x}{\|\theta - \theta^0\|} & \text{if } \theta \ne \theta^0, \\ 0 & \text{if } \theta = \theta^0. \end{cases}$$

Condition (1) is verified in Lemma 4. Next as Hahn (1995, Theorem 3) verified, $\mathcal{G}$ forms a Vapnik-Chervonenkis class of functions and it is thus Donsker by Van der Vaart and Wellner (1996, Theorem 2.6.7 and 2.5.2), which verifies condition (3); note that the assumptions of Theorem 3 in Hahn (1995) are all imposed by Assumptions A.1–A.5. To verify condition (2), note that (29) states $|\rho_\tau(y_i - x_i^T\theta) - \rho_\tau(y_i - x_i^T\theta^0)| \le 2|x_i^T(\theta - \theta^0)|$. Hence, $\sup_{\theta \in U(\theta^0,\delta)} |r(y_i, x_i, \theta)| \le 3\|x_i\|$ has finite second moments ($\sup_{z \in R} |\varphi_\tau(z)| \le 1$ and x_i is bounded by Assumption A.3). The continuity of r and the dominated convergence theorem then imply that $Er^2(y_i, x_i, \theta) \to 0$ as $\theta \to \theta^0$, which verifies condition (2).

Applying Van der Vaart and Wellner (1996, Lemma 3.2.19) leads to the asymptotic linearity $n^{1/2}(\widehat{\theta}_n - \theta^0) = -J^{-1}G_n\{\varphi_\tau(y_i - x_i^T\theta^0)x_i\} + o_p(1)$. The central limit theorem can be applied to $G_n\{\varphi_\tau(y_i - x_i^T\theta^0)x_i\}$ due to Assumptions A.3, which results in the asymptotic normality of $\widehat{\theta}_n$; the form of matrix J is derived in Lemma 4.

To derive the last statement of the theorem, note that the asymptotic linearity given by Lemma 3.2.19 of Van der Vaart and Wellner (1996) is a direct consequence of the fact that the process $G_n\{r(y_i, x_i, \theta)\}$ converges to a Gaussian process on $U(\theta^0, \delta)$. As the differentiability of $E_{P_0}[\rho_\tau(y_i - x_i^T\theta)]$ and condition (30) for P_0 are verified above, the result of Lemma 3.2.19 applies to the triangular structure with sample data of size n generated from P_n if $G_{n,P_n}\{r(y_i, x_i, \theta)\}$ is shown to converge in distribution to a Gaussian process corresponding to P_0 (uniformly on $\mathcal{G}$). This follows from Van der Vaart and Wellner (1996, Theorem 2.8.9) because the assumptions of Theorem 2.8.9 hold: (i) $\mathcal{G}$ is a Vapnik-Chervonenkis class and satisfies thus the uniform entropy condition by Van der Vaart (1996, Theorem 2.6.7), (ii) the envelope of $\mathcal{G}$ given by $\sup_{\theta \in U(\theta^0,\delta)} |r(y_i, x_i, \theta)| \le 3\|x_i\|$ satisfies

$E_{P_n}(3\|x_i\|) < \infty$ and the Lindenberg condition $\limsup_{n\to\infty} E_{P_n}(3\|x_i\| \| \|x_i\| > \epsilon\sqrt{n}) = 0$ for any $\epsilon > 0$ by the compact-support Assumption A.3, and (iii) $\sup_{\theta',\theta''\in U(\theta^0,\delta)} |E_{P_n}\{r(y_i,x_i,\theta') - r(y_i,x_i,\theta'')\}^2 - E_{P_0}\{r(y_i,x_i,\theta') - r(y_i,x_i,\theta'')\}^2| \to 0$ as $n \to \infty$ holds by the assumption of the theorem. □

Lemma 6. *Under Assumptions A.1–A.5 and A.7, $b(\widetilde{F}(\beta^0),\beta^0) \neq b(\widetilde{F}(\beta^1),\beta^1)$ for any $\beta^1 \in B$ such that $\beta^1 \neq \beta^0$.*

Proof: First, note that under the listed assumptions, $b(\widetilde{F}^0,\beta)$ is one-to-one, where $\widetilde{F}^0 = \widetilde{F}(\beta^0)$. Suppose that $b(\widetilde{F}^0,\beta^0) = b(\widetilde{F}(\beta^0),\beta^0) = b(\widetilde{F}(\beta^1),\beta^1)$. Since $b(\widetilde{F}^0,\beta)$ is one-to-one, the vector β^0, which satisfies $b(\widetilde{F}^0,\beta^0) = b(F_y,\beta^0) = \theta^0$ by Theorem 1, uniquely solves the QR moment condition

$$E[F_y(b(\widetilde{F}^0,\beta^0)|x_i) - \tau) \cdot x_i] = 0 \tag{31}$$

(see equation (25)). As $F_y(t|x_i) = 0$ for any $t < 0$ and $P(x_i^T\theta^0 = 0) = 0$ by Assumption A.7, (31) can be written as

$$E[I(x_i^T b(F_y,\beta^0) > 0) \cdot \{F_y(x_i^T b(F_y,\beta^0)|x_i) - \tau\} \cdot x_i] = E[I(x_i^T b(F_y,\beta^0) < 0) \cdot \tau \cdot x_i] = k_0. \tag{32}$$

Next, the QR moment condition for the simulated data $\{\widetilde{y}_i^s(\beta^1), x_i\}_{i=1}^n$ is (see equations (24)–(25))

$$E[(\widetilde{F}_{\widetilde{y}(\beta^1)}(x_i^T b\{\widetilde{F}(\beta^1),\beta^1\}|x_i) - \tau) \cdot x_i] = 0. \tag{33}$$

Because the censored distribution $\widetilde{F}_{\widetilde{y}(\beta^1)}(t|x_i) = 0$ for all $t < 0$, we can again rewrite it as

$$E[I(x_i^T b\{\widetilde{F}(\beta^1),\beta^1\} > 0) \cdot (\widetilde{F}_{\widetilde{y}(\beta^1)}(x_i^T\theta^0|x_i) - \tau) \cdot x_i] = E[I(x_i^T b\{\widetilde{F}(\beta^1),\beta^1\} < 0)\tau \cdot x_i].$$

Recalling that $\theta^0 = b(F_y,\beta^0) = b(\widetilde{F}^0,\beta^0) = b(\widetilde{F}(\beta^0),\beta^0) = b(\widetilde{F}(\beta^1),\beta^1)$ and that $\widetilde{F}_{\widetilde{y}(\beta)}(t|x_i) = \Phi_\tau\{(t - x_i^T\beta)\widetilde{\sigma}^{-1}(x_i,\beta)|x_i\}$ for $t > 0$, (33) becomes

$$E[I(x_i^T b(\widetilde{F}^0,\beta^0) > 0) \cdot (\Phi_\tau\{[x_i^T b(\widetilde{F}(\beta^1),\beta^1) - x_i^T\beta^1] \cdot \widetilde{\sigma}^{-1}(x_i,\beta^1)|x_i\} - \tau) \cdot x_i] = k_0.$$

By substituting (13), where θ^0 is replaced by $b(\widetilde{F}^0,\beta^0)$, we get

$$\begin{aligned} E\Bigg[& I(x_i^T b(\widetilde{F}^0,\beta^0) > 0) I(x_i^T\beta^1 \le 0) \left\{ \Phi_\tau \left(\frac{x_i^T b\{\widetilde{F}(\beta^1),\beta^1\} - x_i^T\beta^1}{x_i^T b(\widetilde{F}^0,\beta^0) - x_i^T\beta^1} \cdot \Phi_\tau^{-1}\{F_y(x_i^T b(\widetilde{F}^0,\beta^0)|x_i)\} \right) - \tau \right\} x_i \\ & + I(x_i^T b(\widetilde{F}^0,\beta^0) > 0) I(x_i^T\beta^1 > 0) \\ & \times \left\{ \Phi_\tau \left(\frac{x_i^T b\{\widetilde{F}(\beta^1),\beta^1\} - x_i^T\beta^1}{x_i^T b(\widetilde{F}^0,\beta^0) - x_i^T\beta^1} \cdot \Phi_\tau^{-1}\{F_y(x_i^T b(\widetilde{F}^0,\beta^0)|x_i) - F_y(x_i^T\beta^1|x_i) + \tau\} \right) - \tau \right\} x_i \Bigg] = k_0. \end{aligned}$$

Recalling again $b(\widetilde{F}^0, \beta^0) = b(\widetilde{F}(\beta^0), \beta^0) = b(\widetilde{F}(\beta^1), \beta^1)$, we obtain

$$\begin{aligned} &E[I(x_i^T b(\widetilde{F}^0, \beta^0) > 0) \cdot I(x_i^T \beta^1 \leq 0) \cdot \{F_y(x_i^T b(\widetilde{F}^0, \beta^0)|x_i) - \tau\} \cdot x_i \\ &\quad + I(x_i^T b(\widetilde{F}^0, \beta^0) > 0) \cdot I(x_i^T \beta^1 > 0) \cdot \{F_y(x_i^T b(\widetilde{F}^0, \beta^0)|x_i) - F_y(x_i^T \beta^1|x_i)\} \cdot x_i] = k_0. \quad (34) \end{aligned}$$

Using identity (32), (34) can be simplified to

$$E[I(x_i^T b(\widetilde{F}^0, \beta^0) > 0) \cdot I(x_i^T \beta^1 > 0) \cdot (F_y(x_i^T \beta^1|x_i) - \tau) \cdot x_i] = 0. \quad (35)$$

Equation (35) is the asymptotic moment condition of CRQ with observations $I(x_i^T b(\widetilde{F}^0, \beta^0) > 0) = I(x_i^T \theta^0 > 0)$. Since J_{crq} is positive definite by Assumption A.5, (35) identifies the true parameter value β^0 (Powell, 1986a) and thus $\beta^0 = \beta^1$. $\square$

Lemma 7. *Under Assumptions A.1–A.5, A.7, and A.8, it holds for $n \to \infty$ that*

$$\sup_{x \in X} \left| \widehat{F}_{y,n}(x_i^T \widehat{\theta}_n|x) - F_y(x_i^T \theta^0|x) \right| = O_p(n^{-\min\{k_1, 1/2\}}).$$

Proof: Let $\epsilon > 0$ and choose $\delta > 0$ such that $P(|x_i^T \theta^0| \leq \delta) < \epsilon/2$ (this is possible due to Assumption A.7). Since $\widehat{\theta}_n$ is $\sqrt{n}$-consistent by Theorem 5 and x_i has a finite support X by Assumption A.3, there is some n_0 such that, for all $n \geq n_0$, $P(x_i^T \theta^0 < -\delta$ and $x_i^T \widehat{\theta}_n \geq -\delta/2) < \epsilon/4$ and $P(x_i^T \theta^0 > \delta$ and $x_i^T \widehat{\theta}_n \leq \delta/2) < \epsilon/4$. Because $\widehat{F}_{y,n}(t|x) = 0$ and $F_y(t|x) = 0$ for any $t < 0$ and $|x_i^T \theta| \leq K$ due to the compactness of Θ and X, we can thus write with an arbitrarily high probability $1 - \epsilon$ that

$$\begin{aligned} \sup_{x \in X} \left| \widehat{F}_{y,n}(x_i^T \widehat{\theta}_n|x) - F_y(x_i^T \theta^0|x) \right| &= I(x_i^T \theta^0 < -\delta) I(x_i^T \widehat{\theta}_n < -\frac{\delta}{2}) \cdot \sup_{x \in X} \left| \widehat{F}_{y,n}(x_i^T \widehat{\theta}_n|x) - F_y(x_i^T \theta^0|x) \right| \\ &+ I(x_i^T \theta^0 > \delta) I(x_i^T \widehat{\theta}_n > \frac{\delta}{2}) \cdot \sup_{x \in X} \left| \widehat{F}_{y,n}(x_i^T \widehat{\theta}_n|x) - F_y(x_i^T \theta^0|x) \right| \\ &= I(x_i^T \theta^0 > \delta) I(x_i^T \widehat{\theta}_n > \frac{\delta}{2}) \\ &\quad \times \sup_{x \in X} \left| \widehat{F}_{y,n}(x_i^T \widehat{\theta}_n|x) - F_y(x_i^T \widehat{\theta}_n|x) + F_y(x_i^T \widehat{\theta}_n|x) - F_y(x_i^T \theta^0|x) \right| \\ &\leq \sup_{x \in X} \sup_{\delta/2 \leq t \leq K} \left| \widehat{F}_{y,n}(t|x) - F_y(t|x) \right| \quad (36) \\ &+ I(x_i^T \theta^0 > \delta) I(x_i^T \widehat{\theta}_n > \frac{\delta}{2}) \sup_{x \in X} \left| F_y(x_i^T \widehat{\theta}_n|x) - F_y(x_i^T \theta^0|x) \right|. \quad (37) \end{aligned}$$

The term (36) behaves as $O_p(n^{-k_1})$ for $n \to \infty$ by Assumption A.8. To bound the other term (37), it can be rewritten using the mean-value theorem as

$$I(x_i^T \theta^0 > \delta) I(x_i^T \widehat{\theta}_n > \frac{\delta}{2}) \sup_{x \in X} \left| f_y(x_i^T \xi_n|x_i) \cdot x_i^T (\widehat{\theta}_n - \theta^0) \right|,$$

where ξ_n represents a linear combination of $\widehat{\theta}_n$ and θ^0. Since the conditional density $f_y(t|x)$ is uniformly bounded for $t > 0$ by Assumption A.4 and the support of x_i is compact by Assumption A.3, the $\sqrt{n}$-consistency of $\widehat{\theta}_n$ implies that (37) behaves as $O_p(n^{-1/2})$ for $n \to \infty$, which concludes the proof. □

Lemma 8. *Under Assumptions A.1–A.5, A.7, and A.8, it holds for any $\beta \in B$, sufficiently small $\delta > 0$, $l \in \{1, 2\}$, and $n \to \infty$ that*

$$|\widehat{\sigma}_n(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)| = o_p(1), \tag{38}$$

$$\sup_{\beta \in U(\beta^0,\delta)} |\widehat{\sigma}_n(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)| = o_p(1), \tag{39}$$

$$E|\widehat{\sigma}_n(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)|^l = o(1), \tag{40}$$

$$E(\sup_{\beta \in U(\beta^0,\delta)} |\widehat{\sigma}_n(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)|)^l = o_p(1). \tag{41}$$

Proof: Note first that Assumption A.7 implies $P(F_y(x_i^T\theta^0|x_i) = \tau) = 0$ and $P(F_y(x_i^T\theta^0|x_i) = F_y(x_i^T\beta|x_i)) = 0$ for any $\beta \in U(\beta^0, \delta)$ and some $\delta > 0$ small enough to guarantee $\theta^0 \notin U(\beta^0, \delta)$. Since Lemma 7 and Assumption A.8 imply $\max\{|\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x) - F_y(x_i^T\theta^0|x)|, |\widehat{F}_{y,n}(x_i^T\beta|x) - F_y(x_i^T\beta|x)|\} = O_p(n^{-k_1}) = o_p(n^{-k_0}) = c_n$ uniformly in $x \in X$, the definitions (13) and (16) apply with an arbitrarily high probability as $n \to \infty$. Moreover, Assumptions A.4 and A.7 guarantee for any $\gamma > 0$ that $P(|F_y(x_i^T\theta^0|x_i) - \tau| < \gamma) \to 0$ and $P(|F_y(x_i^T\theta^0|x_i) - F_y(x_i^T\beta|x_i)| < \gamma) \to 0$ as $\gamma \to 0$ (uniformly in $\beta \in U(\beta^0, \delta)$, where a sufficiently small δ ensures that $\beta \neq \theta^0$). Using the consistency of $\widehat{\theta}_n$ (Theorem 5) and the compact support of x_i (Assumption A.3), we can thus assume with an arbitrarily high probability that $\max\{|F_y(x_i^T\theta|x_i) - \tau|, |F_y(x_i^T\theta|x_i) - F_y(x_i^T\beta|x_i)|\} \geq \gamma/2$ for $\theta = \theta^0$ and $\theta = \widehat{\theta}_n$, for a sufficiently small $\gamma > 0$, and a sufficiently large $n \geq n_0(\gamma) \in N$. By Lemma (7), the same statement also holds for $\widehat{F}_{y,n}$. It also holds uniformly in β if $\beta \in U(\beta^0, \delta)$.

We will now prove that (16) is a (uniformly) consistent estimator of (13). To discuss the estimation of $\widetilde{\sigma}(x_i;\beta)$ defined by (13), we need to consider two cases. First for $x_i^T\beta \leq 0$, we define

$$\widehat{\sigma}_{21,n}(x_i;\beta) = \frac{x_i^T\widehat{\theta}_n - x_i^T\beta}{\Phi_\tau^{-1}\{\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i)\}},$$

$$\widetilde{\sigma}_{21}(x_i;\beta) = \frac{x_i^T(\theta^0 - \beta)}{\Phi_\tau^{-1}\{F_y(x_i^T\theta^0|x_i)\}}.$$

We will discuss the convergence of $\widehat{\sigma}_{21,n}(x_i;\beta)$ to $\widetilde{\sigma}_{21}(x_i;\beta)$ only in the case of $x_i^T\theta^0 > c > 0$: on the one hand, $P(|x_i^T\theta^0| \leq c) \to 0$ as $c \to 0$ by Assumption A.7 and this probability can be made arbitrarily small by letting $c \to 0$; on the other hand, $\widehat{\sigma}_{21,n}(x_i;\beta) = \widetilde{\sigma}_{21}(x_i;\beta) = 0$ if $x_i^T\widehat{\theta}_n < 0$ and

$x_i^T\theta^0 < 0$ and $P(x_i^T\widehat{\theta}_n \geq 0 | x_i^T\theta^0 < -c) \to 0$ uniformly in x_i as $n \to \infty$ due to the consistency of $\widehat{\theta}_n$ and Assumption A.3. Next, note that

$$
\begin{aligned}
|\widehat{\sigma}_{21,n}(x_i;\beta) - \widetilde{\sigma}_{21}(x_i;\beta)| &= \left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{\Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))} - \frac{x_i^T(\theta^0 - \beta)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} \right| \\
&\leq \left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{\Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))} - \frac{x_i^T(\widehat{\theta}_n - \beta)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} \right| \qquad (42) \\
&+ \left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} - \frac{x_i^T(\theta^0 - \beta)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} \right|. \qquad (43)
\end{aligned}
$$

As $\max\{|\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i) - \tau|, |F_y(x_i^T\theta|x_i) - \tau|\} \geq \gamma/2$ with an arbitrarily high probability, the first term (42) can be bounded in probability (using Lemma 7) by

$$
\begin{aligned}
& \left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{\Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))} - \frac{x_i^T(\widehat{\theta}_n - \beta)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} \right| \\
\leq & \left| \frac{x_i^T(\widehat{\theta}_n - \beta)\{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i)) - \Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))\}}{\Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} \right| \\
\leq & \ C_1 \left| \frac{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i)) - \Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))}{\Phi_\tau^{-1}(\tau+\gamma/2)\Phi_\tau^{-1}(\tau+\gamma/2)} \right|, \qquad (44)
\end{aligned}
$$

where the constant $\|x_i^T(\widehat{\theta}_n - \beta)\| \leq C_1$ due to the compactness of the parameter and covariate spaces (Assumptions A.1 and A.3). Since $x_i^T\theta^0 > c$ and $P(x_i^T\widehat{\theta}_n < c/2$ and $x_i^T\theta^0 > c) \to 0$ as $n \to \infty$ due to the consistency of $\widehat{\theta}_n$ and the compactness of the covariate space (see the proof of Lemma 7), the term (44) is negligible in probability by Lemma 7 and the continuous mapping theorem applied to $\Phi_\tau^{-1}(\cdot)$ ($\sup_{i\leq n} F_y(x_i^T\theta^0|x_i) < K_F < 1$ given the infinite support of ε_i). As the bound is independent of β, it holds also uniformly in $\beta \in U(\beta^0, \delta)$.

The second term (43) can be bounded in probability in a similar way:

$$
\begin{aligned}
\left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} - \frac{x_i^T(\theta^0 - \beta)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} \right| &\leq \left| \frac{x_i^T(\widehat{\theta}_n - \theta^0)}{\Phi_\tau^{-1}(F_y(x_i^T\theta^0|x_i))} \right| \\
&\leq C_2 \left| \frac{\widehat{\theta}_n - \theta^0}{\Phi_\tau^{-1}(\tau+\gamma/2)} \right| = o_p(1),
\end{aligned}
$$

where the constant $\|x_i\| \leq C_2$ by the compactness of the support X (Assumption A.3) and the last equality follows from the consistency of $\widehat{\theta}_n$. Note that the bound is again independent of β if

$\beta \in U(\beta^0, \delta)$. Hence, it holds for $n \to \infty$ that

$$|\widetilde{\sigma}_{21,n}(x_i;\beta) - \widehat{\sigma}_{21}(x_i;\beta)| = o_p(1)$$

and

$$\sup_{\beta \in U(\beta^0,\delta)} |\widetilde{\sigma}_{21,n}(x_i;\beta) - \widehat{\sigma}_{21}(x_i;\beta)| = o_p(1).$$

To deal with the other definition of $\widetilde{\sigma}(x_i, \beta)$ used when $x_i^T\beta > 0$, we define first

$$\widehat{\sigma}_{22,n}(x_i;\beta) = \frac{x_i^T\widehat{\theta}_n - x_i^T\beta}{\Phi_\tau^{-1}[\min\{\max\{\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i) - \widehat{F}_{y,n}(x_i^T\beta|x_i) + \tau, 0\}, 1\}]},$$

$$\widetilde{\sigma}_{22}(x_i;\beta) = \frac{x_i^T(\theta^0 - \beta)}{\Phi_\tau^{-1}[\min\{\max\{F_y(x_i^T\theta^0|x_i) - F_y(x_i^T\beta|x_i) + \tau, 0\}, 1\}]}.$$

For notational convenience, let $\widehat{q}_n(x_i,\beta) = \widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i) - \widehat{F}_{y,n}(x_i^T\beta|x_i) + \tau$ and $q(x_i,\beta) = F_y(x_i^T\theta^0|x_i) - F_y(x_i^T\beta|x_i) + \tau$. Further let $\hat{z}_n(x_i;\beta) = \Phi_\tau^{-1}[\min\{\max\{\widehat{q}_n(x_i,\beta), 0\}, 1\}]$ and $z(x_i;\beta) = \Phi_\tau^{-1}[\min\{\max\{q(x_i,\beta), 0\}, 1\}]$. Similarly to the first case, we will discuss the convergence of $\widehat{\sigma}_{22,n}(x_i;\beta)$ to $\widetilde{\sigma}_{22}(x_i;\beta)$ only in the case of $q(x_i,\beta) \in (c, 1-c)$, $c > 0$: on the one hand, $P(q(x_i,\beta) \in (-c, c) \cup (1-c, 1+c)) \to 0$ as $c \to 0$ by Assumptions A.7 and A.9 (uniformly if $\beta \in U(\beta^0,\delta)$) and this probability can be made arbitrarily small letting $c \to 0$; on the other hand, $\widehat{\sigma}_{22,n}(x_i;\beta) = \widetilde{\sigma}_{22}(x_i;\beta) = 0$ if $\widehat{q}_n(x_i,\beta) \notin (0,1)$ and $q(x_i,\beta) \notin (0,1)$ and $P(\widehat{q}_n(x_i,\beta) \in \langle 0,1\rangle$ and $q(x_i,\beta) \notin (-c, 1+c)) \to 0$ as $n \to \infty$ due to the consistency of $\widehat{\theta}_n$ and Lemma 7.

Consequently, the fact that $\max\{|\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i) - \widehat{F}_{y,n}(x_i^T\beta|x_i)|, |F_y(x_i^T\theta|x_i) - F_y(x_i^T\beta|x_i)|\} \geq \gamma/2$ with an arbitrarily high probability implies that $|\widehat{\sigma}_{22,n}(x_i;\beta) - \widetilde{\sigma}_{22}(x_i;\beta)|$ is bounded in probability by

$$\begin{aligned}
|\widehat{\sigma}_{22,n}(x_i;\beta) - \widetilde{\sigma}_{22}(x_i;\beta)| &= \left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{\hat{z}_n(x_i;\beta)} - \frac{x_i^T(\theta^0 - \beta)}{z(x_i;\beta)} \right| \\
&\leq \left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{\hat{z}_n(x_i;\beta)} - \frac{x_i^T(\widehat{\theta}_n - \beta)}{z(x_i;\beta)} \right| + \left| \frac{x_i^T(\widehat{\theta}_n - \beta)}{z(x_i;\beta)} - \frac{x_i^T(\theta^0 - \beta)}{z(x_i;\beta)} \right| \\
&\leq \left| \frac{x_i^T(\widehat{\theta}_n - \beta)\{\hat{z}_n(x_i;\beta) - z(x_i;\beta)\}}{\hat{z}_n(x_i;\beta) z(x_i;\beta)} \right| + \left| \frac{x_i^T(\widehat{\theta}_n - \theta^0)}{z(x_i;\beta)} \right| \\
&\leq C_1 \left| \frac{\hat{z}_n(x_i;\beta) - z(x_i;\beta)}{\Phi_\tau^{-1}(\tau + \gamma/2)\Phi_\tau^{-1}(\tau + \gamma/2)} \right| + C_2 \sup_{i \leq n} \left| \frac{\widehat{\theta}_n - \theta^0}{\Phi_\tau^{-1}(\tau + \gamma/2)} \right|.
\end{aligned}$$

First, note that $|\hat{z}_n(x_i;\beta) - z(x_i;\beta)| = o_p(1)$ as $n \to \infty$ (uniformly in $\beta \in U(\beta^0,\delta)$) since $|\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i) -$

$F_y(x_i^T\theta^0|x_i)| = o_p(1)$ by Lemma 7, $|\widehat{F}_{y,n}(x_i^T\beta|x_i) - F_y(x_i^T\beta|x_i)| = o_p(1)$ by Assumption A.8, and the claim follows by the continuous mapping theorem. As $\widehat{\theta}_n$ is consistent by Theorem 5, we can conclude that $|\widehat{\sigma}_{22,n}(x_i;\beta) - \widetilde{\sigma}_{22}(x_i;\beta)| = o_p(1)$. As the bounds are again valid independently of $\beta \in U(\beta^0, \delta)$, we have also proved that

$$\sup_{\beta\in U(\beta^0,\delta)} |\widetilde{\sigma}_{22}(x_i;\beta) - \widehat{\sigma}_{22,n}(x_i;\beta)| = o_p(1). \tag{45}$$

To derive (40)–(41), let $\{c_n\}_{n=1}^{\infty}$ be a sequence such that $c_n = O(n^{-k_0})$ defining $\widehat{\sigma}_n(x_i;\beta)$ and $a_n = \Phi_\tau^{-1}(\tau + c_n)$. Given the established convergence in probability, $a_n|\widehat{\sigma}_n(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)| = o_p(a_n)$ and $\sup_{\beta\in U(\beta^0,\delta)} a_n|\widehat{\sigma}_n(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)| = o_p(a_n)$, the last two claims of the lemma follow if $\{a_n\cdot\widehat{\sigma}_{21,n}(x_i;\beta)\}^2$ and $\{a_n\cdot\widehat{\sigma}_{22,n}(x_i;\beta)\}^2$ are uniformly integrable (Davidson, 1994, Theorem 18.14). Using Davidson (1994, Theorem 12.10), we will prove this by showing that the $(2+\epsilon)$th moment $\{a_n\cdot\widehat{\sigma}_{21,n}(x_i;\beta)\}^{2+\epsilon}$ is uniformly bounded in n and β for some $\epsilon > 0$ ($\{a_n\cdot\widehat{\sigma}_{22,n}(x_i;\beta)\}^{2+\epsilon}$ can be bounded analogously). By the definition of $\widehat{\sigma}_{21,n}(x_i;\beta)$, we have

$$a_n\cdot\widehat{\sigma}_{21,n}(x_i;\beta) = a_n\cdot\frac{x_i^T\widehat{\theta}_n - x_i^T\beta}{\Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i))}.$$

Since $|\Phi_\tau^{-1}\{\widehat{F}_{y,n}(x_i^T\widehat{\theta}_n|x_i)\}|$ is bounded from below by $|a_n|$ by the definition of $\widehat{\sigma}_n(x_i,\beta)$, it follows for some $\epsilon > 0$ and some $K_\epsilon > 0$

$$|a_n|^{2+\epsilon}\cdot|\widehat{\sigma}_{21,n}(x_i;\beta)|^{2+\epsilon} \le |a_n|^{2+\epsilon}\cdot\frac{|x_i^T(\widehat{\theta}_n-\beta)|^{2+\epsilon}}{|\Phi_\tau^{-1}(|a_n|)|^{2+\epsilon}} \le K_\epsilon|x_i^T(\widehat{\theta}_n-\beta)|^{2+\epsilon}.$$

Since the parameter spaces B and Θ are compact by Assumption A.1 and x_i has a compact support by Assumption A.3, $|x_i^T(\widehat{\theta}_n-\beta)|^{2+\epsilon}$ is bounded uniformly in $n \in N$ and in $\beta \in B$ and hence $\{a_n\cdot\widehat{\sigma}_{21}(x_i;\beta)\}^2$ is uniformly integrable. Similarly, uniform integrability can be established for $\{a_n\cdot\widehat{\sigma}_{22,n}(x_i;\beta)\}^2$. Thus, the convergence-in-mean claims follows from the convergence in probability: for example, $a_n|\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)| = o_p(a_n)$ implies $E\{a_n|\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)|\}^l = o_p(a_n^l)$ for $l \in \{1,2\}$ as $n\to\infty$. The claim of the lemma follows after standardizing by a_n^l. □

Corollary 9. *Under Assumptions A.1–A.5, A.7, and A.8, it holds for some $\delta > 0$ and $n\to\infty$:*

$$\sup_{\beta\in U(\beta^0,\delta)}\|\widehat{F}(\beta)-\widetilde{F}(\beta)\|_\infty = \sup_{\beta\in U(\beta^0,\delta)}\sup_{t\in R}\left|\Phi_\tau\left[\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)}\right]-\Phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]\right| = o_p(1),$$

$$\sup_{\beta\in U(\beta^0,\delta)}\sup_{t\in R}\left|\frac{1}{\widehat{\sigma}_n(x_i;\beta)}\phi_\tau\left[\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)}\right]-\frac{1}{\widetilde{\sigma}(x_i;\beta)}\phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]\right| = o_p(1).$$

Further, function $\widetilde{F}(\beta)$ is continuous in β on $U(\beta^0,\delta)$:

$$\lim_{\beta'\to\beta}\|\widetilde{F}(\beta')-\widetilde{F}(\beta)\|_\infty=\lim_{\beta'\to\beta}\sup_t\left|\Phi_\tau\left[\frac{t-x_i^T\beta'}{\widetilde{\sigma}(x_i;\beta')}\right]-\Phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]\right|=o_p(1).$$

Proof: First, note that the continuity of $\widetilde{\sigma}(x_i;\beta)$ in β implies that there exists $\delta>0$ such that $\widetilde{\sigma}(x_i;\beta)>C_\sigma/2$ for any x_i and all $\beta\in U(\beta^0,\delta)$. At the same time, $\widetilde{\sigma}(x_i;\beta)<K_\sigma$ for any $\beta\in U(\beta^0,\delta)$ as explained at the start of the appendix. Hence, the mean value theorem implies that

$$\begin{aligned}\left|\Phi_\tau\left[\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)}\right]-\Phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]\right| &\leq \phi_\tau(\xi_n)\left|\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)}-\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right|\\ &= \phi_\tau(\xi_n)\left|\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)\widetilde{\sigma}(x_i;\beta)}\right|\cdot|\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)|,\end{aligned}$$

where $\xi_n=\xi_n^1=(t-x_i^T\beta)/\widehat{\sigma}_n(x_i;\beta)$ or $\xi_n=\xi_n^2=(t-x_i^T\beta)/\widetilde{\sigma}(x_i;\beta)$ depending on which of $\phi_\tau(\xi_n^1)$ and $\phi_\tau(\xi_n^2)$ is larger. Therefore, $\phi_\tau(\xi_n)\left|t-x_i^T\beta\right|/\widehat{\sigma}_n(x_i;\beta)\widetilde{\sigma}(x_i;\beta)\leq 2/C_\sigma$ and the first claim follows from Lemma 8.

Similarly, we can apply the mean value theorem to obtain

$$\begin{aligned}&\left|\frac{1}{\widehat{\sigma}_n(x_i;\beta)}\phi_\tau\left[\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)}\right]-\frac{1}{\widetilde{\sigma}(x_i;\beta)}\phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]\right|\\ \leq\ &\frac{1}{\widehat{\sigma}_n(x_i;\beta)\widetilde{\sigma}(x_i;\beta)}\cdot\left|\phi_\tau\left[\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)}\right]-\phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]\right|\widetilde{\sigma}(x_i;\beta)\\ +\ &\frac{1}{\widehat{\sigma}_n(x_i;\beta)\widetilde{\sigma}(x_i;\beta)}\cdot\phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]|\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)|\\ \leq\ &\frac{\widetilde{\sigma}(x_i;\beta)}{\widehat{\sigma}_n(x_i;\beta)\widetilde{\sigma}(x_i;\beta)}\cdot\phi_\tau'(\xi_n)\left|\frac{t-x_i^T\beta}{\widehat{\sigma}_n(x_i;\beta)\widetilde{\sigma}(x_i;\beta)}\right|\cdot|\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)|\\ +\ &\frac{1}{\widehat{\sigma}_n(x_i;\beta)\widetilde{\sigma}(x_i;\beta)}\cdot\phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]|\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)|,\end{aligned}$$

where again $\xi_n=\xi_n^1=(t-x_i^T\beta)/\widehat{\sigma}_n(x_i;\beta)$ or $\xi_n=\xi_n^2=(t-x_i^T\beta)/\widetilde{\sigma}(x_i;\beta)$. As $\phi_\tau'(t)=-t\phi_\tau(t)$, $t^2\phi_\tau(t)$ is uniformly bounded, and the variance functions are within interval $(C_\sigma/4,4K_\sigma)$ with a probability arbitrarily close to 1, $P(\widehat{\sigma}_n(x_i;\beta)\in(C_\sigma/4,4K_\sigma))\to 1$ as $n\to\infty$, the second claim of the lemma follows again from Lemma 8.

Finally, the continuity in β can be verified in the following way:

$$\begin{aligned}
\left|\Phi_\tau\left[\frac{t-x_i^T\beta'}{\widetilde{\sigma}(x_i;\beta')}\right]-\Phi_\tau\left[\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right]\right| &\leq \phi_\tau(\xi_n)\left|\frac{t-x_i^T\beta'}{\widetilde{\sigma}(x_i;\beta')}-\frac{t-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta)}\right| \\
&= \phi_\tau(\xi_n)\left|\frac{(t-x_i^T\beta')\widetilde{\sigma}(x_i;\beta)-(t-x_i^T\beta)\widetilde{\sigma}(x_i;\beta')}{\widetilde{\sigma}(x_i;\beta')\widetilde{\sigma}(x_i;\beta)}\right| \\
&\leq \phi_\tau(\xi_n)\left|\frac{(t-x_i^T\beta')}{\widetilde{\sigma}(x_i;\beta')\widetilde{\sigma}(x_i;\beta)}\right|\cdot|\widetilde{\sigma}(x_i;\beta')-\widetilde{\sigma}(x_i;\beta)| \\
&+ \phi_\tau(\xi_n)\left|\frac{x_i^T\beta'-x_i^T\beta}{\widetilde{\sigma}(x_i;\beta')}\right|,
\end{aligned}$$

where $\xi_n = (t-x_i^T\beta)/\widetilde{\sigma}(x_i;\beta)$ or $\xi_n = (t-x_i^T\beta')/\widetilde{\sigma}(x_i;\beta')$, whichever leads to a higher value of $\phi_\tau(\xi_n)$. The last claim of the corollary is then implied by the continuity of $\widetilde{\sigma}(x_i;\beta)$ is β, uniform boundedness of $\widetilde{\sigma}(x_i;\beta) > C_\sigma/2$ for any x_i and all $\beta \in U(\beta^0,\delta)$, and the boundedness of X. $\square$

Lemma 10. *Under Assumptions A.1–A.5, A.7, and A.8, it holds for any $\beta \in B$, $1 \leq s \leq S$, and $n \to \infty$ that*

$$E\{\widehat{y}_i^s(\beta)-\widetilde{y}_i^s(\beta)\}^2 = o(1), \tag{46}$$

and for any $\beta \in B$, $\theta \in \Theta$, and $n \to \infty$, that

$$E\left|I(\widehat{y}_i^s(\beta) \leq x_i^T\theta) - I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\right|^2 = o(1). \tag{47}$$

Furthermore, statements (46) and (47) along with

$$E\left|I(\widehat{y}_i^s(\beta') \leq x_i^T\theta')I(\widehat{y}_i^s(\beta) \leq x_i^T\theta) - I(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\right|^2 = o(1), \tag{48}$$

$$E\left|\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta')I(\widehat{y}_i^s(\beta) \leq x_i^T\theta) - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\right|^2 = o(1), \tag{49}$$

and

$$E\left|\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta')\rho_\tau(\widehat{y}_i^s(\beta) \leq x_i^T\theta) - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')\rho_\tau(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\right|^2 = o(1) \tag{50}$$

hold also uniformly with respect to $\beta, \beta' \in U(\beta^0,\delta)$ and $\theta, \theta' \in U(\theta^0,\delta)$ for a sufficiently small $\delta > 0$.

Proof: Regarding the first claim, we consider first latent variables written in the form $\widetilde{y}_i^{s*}(\beta) = x_i^T\beta + \nu_i^s \cdot \widetilde{\sigma}(x_i;\beta)$ and $\widehat{y}_i^{s*}(\beta) = x_i^T\beta + \nu_i^s \cdot \widehat{\sigma}_i(x_i;\beta)$, where $\nu_i^s \sim N(\mu_\tau, 1)$. Since $\widetilde{y}_i^{s*}(\beta) - \widehat{y}_i^{s*}(\beta) = \nu_i^s \cdot \{\widehat{\sigma}_i(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)\}$, we obtain from Lemma 8 that

$$\begin{aligned}
E\{\widehat{y}_i^{s*}(\beta)-\widetilde{y}_i^{s*}(\beta)\}^2 &= E[(\nu_i^s)^2\cdot\{\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)\}^2] \\
&= E(\nu_i^s)^2\cdot E[\{\widehat{\sigma}_n(x_i;\beta)-\widetilde{\sigma}(x_i;\beta)\}^2] \\
&= K\cdot o(1),
\end{aligned}$$

where $E(\nu_i^s)^2 < K$ for some $K > 0$ as ν_i^s are independent and identically distributed. The claim (38) follows from $[\widetilde{y}_i^s(\beta) - \widehat{y}_i^s(\beta)]^2 = [\max\{0, \widetilde{y}_i^{s*}(\beta)\} - \max\{0, \widehat{y}_i^{s*}(\beta)\}]^2 \leq [\widetilde{y}_i^{s*}(\beta) - \widehat{y}_i^{s*}(\beta)]^2$. Moreover,

(38) holds uniformly in $\beta \in U(\beta^0, \delta)$ for some $\delta > 0$ since $E[\{\widehat{\sigma}_n(x_i;\beta) - \widetilde{\sigma}(x_i;\beta)\}^2]$ converges uniformly to zero by Lemma 8.

To prove the second claim, consider some given θ, β, and $\epsilon > 0$ and note that $E|I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta) - I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)| = EP(I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta) \neq I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)|x_i, \widehat{\sigma}_n)$, where ν_i^s is independent of x_i and $\widehat{\sigma}_n$. Here, $I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta) \neq I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)$ can occur only if $\widetilde{\sigma}(x_i;\beta)\nu_i^s \leq x_i^T\theta - x_i^T\beta$ and $\widehat{\sigma}_n(x_i;\beta)\nu_i^s > x_i^T\theta - x_i^T\beta$ or vice versa. As the conditional probability is obviously non-zero only if $\widetilde{\sigma}(x_i;\beta) > 0$ or $\widehat{\sigma}_n(x_i;\beta) > 0$, assume $\widetilde{\sigma}(x_i;\beta) > 0$ without loss of generality almost surely (note that, as argued in Corollary 9, the continuity of $\widetilde{\sigma}(x_i;\beta)$ in β implies that there exists $\delta > 0$ such that $\widetilde{\sigma}(x_i;\beta) > C_\sigma/2$ for any x_i and all $\beta \in U(\beta^0, \delta)$). Consider all values of x_i such that $\widetilde{\sigma}(x_i;\beta) > c > 0$, which holds with a probability larger than $1 - \epsilon/2$ for a sufficiently small c. Consequently, $\min\{\widetilde{\sigma}(x_i;\beta), \widehat{\sigma}_n(x_i;\beta)\} > c/2$ holds with a probability larger than $1 - \epsilon$ for a sufficiently large n by Lemma 8. The indicators can then differ only if $\nu_i^s \leq \widetilde{\sigma}(x_i;\beta)^{-1}(x_i^T\theta - x_i^T\beta)$ and $\nu_i^s > \widehat{\sigma}_n(x_i;\beta)^{-1}(x_i^T\theta - x_i^T\beta)$ or vice versa. Hence, it holds with probability at least $1 - \epsilon$ that

$$
\begin{aligned}
& P(I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta) \neq I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)|x_i, \widehat{\sigma}_n) \\
\leq \quad & P(I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta) \neq I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)|x_i, \widehat{\sigma}_n) \\
\leq \quad & |\Phi_\tau\{\widetilde{\sigma}(x_i;\beta)^{-1}(x_i^T\theta - x_i^T\beta)\} - \Phi_\tau\{\widehat{\sigma}_n(x_i;\beta)^{-1}(x_i^T\theta - x_i^T\beta)\}| \\
\leq \quad & \phi_\tau\{\xi_n\} \cdot |\widetilde{\sigma}(x_i;\beta)^{-1} - \widehat{\sigma}_n(x_i;\beta)^{-1}| \cdot |x_i^T\theta - x_i^T\beta| \\
\leq \quad & \phi_\tau\{\xi_n\} \cdot (c/2)^{-2}|\widetilde{\sigma}(x_i;\beta) - \widehat{\sigma}_n(x_i;\beta)| \cdot |x_i^T\theta - x_i^T\beta|,
\end{aligned}
$$

where ξ_n lies between $\widetilde{\sigma}(x_i;\beta)^{-1}(x_i^T\theta - x_i^T\beta)$ and $\widehat{\sigma}_n(x_i;\beta)^{-1}(x_i^T\theta - x_i^T\beta)$. Since $P(I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta) \neq I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)|x_i, \widehat{\sigma}_n) \leq 1$ for any x_i and $\widehat{\sigma}_n$, the boundedness of the normal density ϕ_τ by some $K > 0$ and Lemma 8 then imply that

$$
\begin{aligned}
E|I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta) - I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)| \quad & \leq \quad K\frac{4}{c^2}E[|\widetilde{\sigma}(x_i;\beta) - \widehat{\sigma}_n(x_i;\beta)| \cdot |x_i^T\theta - x_i^T\beta|] + \epsilon \qquad (51) \\
& = \quad o(1) + \epsilon
\end{aligned}
$$

because X, Θ, and B are bounded. Letting $\epsilon \to 0$ completes the proof. As $\widetilde{\sigma}(x_i;\beta) > C_\sigma/2$ for any x_i and all $\beta \in U(\beta^0, \delta)$, the uniform convergence of $\widehat{\sigma}_n(x_i;\beta)$ to $\widetilde{\sigma}(x_i;\beta)$ in Lemma 8 and the boundedness of the parameter spaces in Assumption A.1 and A.3 again imply that the upper bound (51) converges to zero uniformly in $\beta \in U(\beta^0, \delta)$ and $\theta \in U(\theta^0, \delta)$.

Next, as any indicator or difference of two indicators is smaller or equal to 1 in absolute value, (48) follows directly from claim (47) by noting that

$$
\begin{aligned}
& E\left|I(\widehat{y}_i^s(\beta) \leq x_i^T\theta)I(\widehat{y}_i^s(\beta') \leq x_i^T\theta') - I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)I(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')\right|^2 \\
\leq \quad & E\left|\{I(\widehat{y}_i^s(\beta) \leq x_i^T\theta) - I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\}I(\widehat{y}_i^s(\beta') \leq x_i^T\theta')\right|^2 \\
+ \quad & E\left|I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\{I(\widehat{y}_i^s(\beta') \leq x_i^T\theta') - I(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')\}\right|^2 \\
+ \quad & 2E\left[\left|\{I(\widehat{y}_i^s(\beta) \leq x_i^T\theta) - I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\}I(\widehat{y}_i^s(\beta') \leq x_i^T\theta')\right|\right. \\
& \qquad \left.\times\left|I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\{I(\widehat{y}_i^s(\beta') \leq x_i^T\theta') - I(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')\}\right|\right].
\end{aligned}
$$

Finally, (49) has to be verified ((50) can be verified in the same way). Writing

$$
\begin{aligned}
& E\left|\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta')I(\widehat{y}_i(\beta) \leq x_i^T\theta) - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\right|^2 \\
\leq \quad & E\left|\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta')\{I(\widehat{y}_i(\beta) \leq x_i^T\theta) - I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\}\right|^2 && (52) \\
+ \quad & E\left|\{\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta') - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')\}I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\right|^2 && (53) \\
+ \quad & 2E\left[\left|\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta')\{I(\widehat{y}_i(\beta) \leq x_i^T\theta) - I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\}\right|\right. && (54) \\
& \qquad \left.\times\left|\{\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta') - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')\}I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\right|\right],
\end{aligned}
$$

we can bound the expressions (52)–(54) in the following way. Term (52) can be bounded using the Cauchy-Schwarz inequality by

$$
E\left|\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta')\right|^2 E\left|\{I(\widehat{y}_i(\beta) \leq x_i^T\theta) - I(\widetilde{y}_i^s(\beta) \leq x_i^T\theta)\}\right|^2
$$

and its uniform convergence to 0 follows from claim (47). Similarly to (29), term (53) is bounded by

$$
E\left|\{\rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T\theta') - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T\theta')\}\right|^2 \leq 2E|\widehat{y}_i^s(\beta') - \widetilde{y}_i^s(\beta')| \tag{55}
$$

and its uniform convergence to 0 follows from claim (46). Term (54) can be dealt with similarly (i.e., using the Cauchy-Schwarz inequality along with (55)), which concludes the proof. □

Lemma 11. *Under Assumptions A.1–A.5 and A.6–A.8, it holds for any $\beta \in B$, $1 \leq s \leq S$, and $n \to \infty$ that*

$$
|\widehat{\theta}_n^s(\beta) - b(\widetilde{F}(\beta), \beta)| = o_p(1),
$$

and in particular, $|\widehat{\theta}_n^s(\beta^0) - \theta^0| = o_p(1)$.

Proof: Given some $\beta \in B$, denote the QR objective functions $\widehat{Q}_n^s(\theta, \beta) = E_n[\rho_\tau(\widehat{y}_i(\beta) - x_i^\top\theta) -$

$\rho_\tau(\widehat{y}_i(\beta) - x_i^\top\theta^0)]$ and $\widetilde{Q}_n^s(\theta,\beta) = E_n[\rho_\tau(\widetilde{y}_i(\beta) - x_i^\top\theta) - \rho_\tau(\widetilde{y}_i(\beta) - x_i^\top\theta^0)]$. Theorem 5 implies that $\widetilde{Q}_n^s(\theta,\beta) \xrightarrow{P} Q_\infty(\theta,\beta)$ for all $\theta \in \Theta$, where $Q_\infty(\theta,\beta)$ is minimized at $b(\widetilde{F}(\beta),\beta)$ (see definition (5)), and because of convexity of $\widetilde{Q}_n^s(\theta,\beta)$, that $\widetilde{\theta}_n^s(\beta) \to b(\widetilde{F}(\beta),\beta)$ in probability. Recall that $b(\widetilde{F}(\beta^0),\beta^0) = \theta^0$ by (10) and Theorem 1. To prove the claim of the lemma, we therefore have to prove that the instrumental criterion $\widehat{Q}_n^s(\theta,\beta)$, which is also convex in θ, has the same pointwise limit as $\widetilde{Q}_n^s(\theta,\beta)$ for all $\theta \in \Theta$ (see Newey and McFadden, 1994, Theorem 2.7).

As $\widehat{Q}_n^s(\theta,\beta) = E_n[\rho_\tau(\widehat{y}_i(\beta) - x_i^\top\theta) - \rho_\tau(\widehat{y}_i(\beta) - x_i^\top\theta^0)]$, we prove the convergence only for the first term $E_n[\rho_\tau(\widehat{y}_i(\beta) - x_i^\top\theta)]$ as the second one is a special case of the first one. We can write

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n \{\tau - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\}(\widehat{y}_i^s(\beta) - x_i^T\theta) \\
= \; & \frac{1}{n}\sum_{i=1}^n \{\tau - I(\widetilde{y}_i^s(\beta) \le x_i^T\theta) + I(\widetilde{y}_i^s(\beta) \le x_i^T\theta) - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\} \\
& \qquad \times\{\widetilde{y}_i^s(\beta) - x_i^T\theta - \widetilde{y}_i^s(\beta) + \widehat{y}_i^s(\beta)\} \\
= \; & E_n[\rho_\tau(\widetilde{y}_i(\beta) - x_i^\top\theta)] + \frac{1}{n}\sum_{i=1}^n \{\tau - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\}\{\widehat{y}_i^s(\beta) - \widetilde{y}_i^s(\beta)\} \qquad (56) \\
& + \frac{1}{n}\sum_{i=1}^n \{I(\widetilde{y}_i^s(\beta) \le x_i^T\theta) - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\}\{\widetilde{y}_i^s(\beta) - x_i^T\theta\}. \qquad (57)
\end{aligned}
$$

Denoting $T_1 = n^{-1}\sum_{i=1}^n \{\tau - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\}\{\widehat{y}_i^s(\beta) - \widetilde{y}_i^s(\beta)\}$ and $T_2 = n^{-1}\sum_{i=1}^n \{I(\widetilde{y}_i^s(\beta) \le x_i^T\theta) - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\}\{\widehat{y}_i^s(\beta) - x_i^T\theta\}$, we have to prove that

$$
|E_n[\rho_\tau(\widehat{y}_i(\beta) - x_i^\top\theta)] - E_n[\rho_\tau(\widetilde{y}_i(\beta) - x_i^\top\theta)]| \le 2|T_1| + 2|T_2| = o_p(1). \qquad (58)
$$

First, consider term T_1. By the Cauchy-Schwarz inequality, we have

$$
\frac{1}{n}\sum_{i=1}^n \{\tau - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\}\{\widehat{y}_i^s(\beta) - \widetilde{y}_i^s(\beta)\} \le \sqrt{\frac{1}{n}\sum_{i=1}^n \{\tau - I(\widehat{y}_i^s(\beta) \le x_i^T\theta)\}^2}\sqrt{\frac{1}{n}\sum_{i=1}^n \{\widehat{y}_i^s(\beta) - \widetilde{y}_i^s(\beta)\}^2}. \qquad (59)
$$

Note that the first term on the right-hand-side of (59) is bounded by 1 irrespective of θ. For the second term, which is independent of θ, we obtain by the Markov inequality

$$\begin{aligned} P\left[\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i^s(\beta)-\widetilde{y}_i^s(\beta))^2>\epsilon\right] &\leq \frac{1}{n\epsilon}E\left[\sum_{i=1}^{n}(\widehat{y}_i^s(\beta)-\widetilde{y}_i^s(\beta))^2\right] \\ &= \frac{1}{n\epsilon}\sum_{i=1}^{n}E[(\widehat{y}_i^s(\beta)-\widetilde{y}_i^s(\beta))^2] \\ &= \frac{1}{\epsilon}E[(\widehat{y}_i^s(\beta)-\widetilde{y}_i^s(\beta))^2]=o(1) \end{aligned}$$

as $n\to\infty$, where the last equality follows from Lemma 10. Hence, $|T_1|=o_p(1)$.

Next, consider term T_2:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\{I(\widetilde{y}_i^s(\beta) &\leq x_i^T\theta)-I(\widehat{y}_i^s(\beta)\leq x_i^T\theta)\}\{\widehat{y}_i^s(\beta)-x_i^T\theta)\} \\ &\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i^s(\beta)-x_i^T\theta))^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\{I(\widetilde{y}_i^s(\beta)\leq x_i^T\theta)-I(\widehat{y}_i^s(\beta)\leq x_i^T\theta)\}^2}. \end{aligned}$$

By applying the Markov inequality and Lemma 10 on the second term on the right-hand-side of the inequality, we again find $|T_2|=o_p(1)$ since X, Θ, and B are compact by Assumptions A.1 and A.3. Thus, (58) holds and it follows that $|\widehat{\theta}_n^s(\beta)-b(\widetilde{F}(\beta),\beta)|=o_p(1)$. □

Appendix C. Proofs of the main asymptotic properties

The proofs in this section rely on the notation introduced at the beginning of Appendix B.

Proof of Theorem 2: First, we show that β^0 is identified. By definition (5), the instrumental criterion yields θ^0 at the true value of the parameter β^0, $b(F_y,\beta^0)=\theta^0$. Theorem 1 and the construction of $\widetilde{F}$ in (10) then imply $b(\widetilde{F}(\beta^0),\beta^0)=b(F_y,\beta^0)=\theta^0$. On the other hand, Lemma 6 indicates that, for any $\beta^1\neq\beta^0$, $\beta^1\in B$, the QR yields different estimates: $b(\widetilde{F}(\beta^1),\beta^1)\neq b(\widetilde{F}(\beta^0),\beta^0)=\theta^0$.

To prove consistency, note that $\widehat{\theta}_n\to\theta^0=b(F_y,\beta^0)$ by Theorem 5. Similarly for any $s=1,\dots,S$, $\widetilde{\theta}_n^s(\beta^0)\to b(\widetilde{F}(\beta^0),\beta^0)=b(F_y,\beta^0)=\theta^0$ and $\widetilde{\theta}_n^s(\beta)\to b(\widetilde{F}(\beta),\beta)\neq\theta^0$ for $\beta\neq\beta^0$ since, for a given $\beta\in B$, the data $(\widetilde{y}_i^s(\beta),x_i)_{i=1}^n$ also satisfy the assumptions of Theorem 5. The same holds also for $\sum_{s=1}^S\widetilde{\theta}_n^s(\beta^0)/S$ as the limits of $\widetilde{\theta}_n^s(\beta)$ are independent of s and S is finite. The III criterion (15) is thus a strictly convex function in $\widehat{\theta}_n$ and $\sum_{s=1}^S\widetilde{\theta}_n^s(\beta^0)/S$, which converges for any β to

$$\arg\min_{\beta\in B}\left[b(F_y,\beta^0)-b(\widetilde{F}(\beta),\beta)\right]^T\Omega\left[b(F_y,\beta^0)-b(\widetilde{F}(\beta),\beta)\right]. \tag{60}$$

Hence by Assumption A.6, any minimizer $\widehat{\beta}_n^{III}$ of (15) satisfies $\sum_{s=1}^S\widetilde{\theta}_n^s(\widehat{\beta}_n^{III})/S\to\theta^0$ in probability as $n \to \infty$ (Newey and McFadden, 1994, Theorem 2.7). As the link function b is one-to-one continuous mapping (Assumption A.6) and the parameter space B is compact (Assumption A.1), $\widehat{\beta}_n^{III}$ has to converge in probability to β^0, which is the unique minimum of (60) (cf. the proof of Theorem 1 in Gourieroux and Monfort, 1993).

The proof for asymptotic normality of $\widehat{\beta}_n^{III}$ is similar to the proof of Proposition 3 in Gourieroux and Monfort (1993), which is however given for a twice continuously differentiable instrumental criterion. By taking the first-order condition of the optimization problem (15) with respect to β, the first-order condition is obtained:

$$\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\widetilde{\theta}_n^s(\widehat{\beta}_n^{III})^T}{\partial\beta}\right]\Omega\left[\widehat{\theta}_n-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\widehat{\beta}_n^{III})\right]=0.$$

Applying the Taylor expansion around β^0 to $\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\widehat{\beta}_n^{III})$, we obtain analogously to Gourieroux and Monfort (1993, equation (51)) for some linear combination ξ_n of β^0 and $\widehat{\beta}_n^{III}$ and for $n \to \infty$ that

$$\left[\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\widetilde{\theta}_n^s(\widehat{\beta}_n^{III})^T}{\partial\beta}\right]\Omega\left[\left\{\widehat{\theta}_n-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta^0)\right\}-\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\widetilde{\theta}_n^s(\xi_n)^T}{\partial\beta}\left\{\widehat{\beta}_n^{III}-\beta^0\right\}\right]=0, \tag{61}$$

and due to the full-rank Assumption A.6, that

$$n^{1/2}(\widehat{\beta}_n^{III}-\beta^0) = \left(\frac{\partial b^T}{\partial\beta}(\widetilde{F}(\beta^0),\beta^0)\Omega\frac{\partial b}{\partial\beta^T}(\widetilde{F}(\beta^0),\beta^0)\right)^{-1} \tag{62}$$

$$\times\frac{\partial b^T}{\partial\beta}(\widetilde{F}(\beta^0),\beta^0)\Omega n^{1/2}\left[\widehat{\theta}_n-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta^0)\right]+o_p(1), \tag{63}$$

where we used $\partial\widetilde{\theta}_n^s(\beta)/\partial\beta^T \to \partial b(\widetilde{F}(\beta),\beta)/\partial\beta^T$ and $\partial b(\widetilde{F}(\widehat{\beta}_n^{III}),\widehat{\beta}_n^{III})/\partial\beta^T \to \partial b(\widetilde{F}(\beta^0),\beta^0)/\partial\beta^T = D$ as $n \to \infty$ in probability (the required continuity of the derivative of the link function and the full rank of its derivative follow from Assumption A.6, whereas the continuity of $\widetilde{F}$ in β follows from its Corollary 9).

Next, by Theorem 5, we have

$$n^{1/2}(\widehat{\theta}_n-\theta^0) = -J^{-1}G_n[\varphi_\tau(y_i - x_i^T\theta^0)x_i]+o_p(1). \tag{64}$$

Since $\{\widetilde{y}_i^s(\beta^0),x_i\}_{i=1}^n$ and $\widetilde{\varepsilon}_i^s \sim N(\mu_\tau\cdot\widetilde{\sigma}(x_i;\beta),\widetilde{\sigma}(x_i;\beta))$ satisfy Assumptions A.1–A.5 and A.7, Theorem 5 also implies

$$n^{1/2}(\widetilde{\theta}_n^s(\beta^0) - \theta^0) = -\widetilde{J}^{-1}G_n[\varphi_\tau(\widetilde{y}_i^s(\beta^0) - x_i^T\theta^0)x_i] + o_p(1) \tag{65}$$

for any simulated path $s = 1, \ldots, S$ and $n \to \infty$. Thus, combining (64) and (65) yields

$$\begin{aligned} n^{1/2}[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta^0)] &= -J^{-1}G_n[\varphi_\tau(y_i - x_i^T\theta^0)x_i] + \widetilde{J}^{-1}\frac{1}{S}\sum_{s=1}^{S}G_n[\varphi_\tau(\widetilde{y}_i^s(\beta^0) - x_i^T\theta^0)x_i] + o_p(1) \\ &= [-J^{-1}, \frac{\widetilde{J}^{-1}}{S}, \ldots, \frac{\widetilde{J}^{-1}}{S}]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix} \varphi_\tau(y_i - x_i^T\theta^0) \\ \varphi_\tau(\widetilde{y}_i^1(\beta^0) - x_i^T\theta^0) \\ \vdots \\ \varphi_\tau(\widetilde{y}_i^S(\beta^0) - x_i^T\theta^0) \end{bmatrix} \otimes x_i + o_p(1). \end{aligned} \tag{66}$$

The random variables $\varphi_\tau(y_i - x_i^T\theta^0)x_i$ and $\varphi_\tau(\widetilde{y}_i^s(\beta^0) - x_i^T\theta^0)x_i$ have zero means (by the definition of θ^0) and finite variances and covariances (due to Assumption A.3), which are computed below. By the central limit theorem, the random vector in (66) thus converges in distribution to a normally distributed random vector. Using the notation Σ and $\widetilde{\Sigma}$ from Assumption A.5 and denoting $\widetilde{K}_{0s} = cov\{\varphi_\tau(y_i - x_i^T\theta^0)x_i, \varphi_\tau(\widetilde{y}_i^s(\beta^0) - x_i^T\theta^0)x_i\}$ and $\widetilde{K}_{rs} = cov\{\varphi_\tau(\widetilde{y}_i^r(\beta^0) - x_i^T\theta^0)x_i, \varphi_\tau(\widetilde{y}_i^s(\beta^0) - x_i^T\theta^0)x_i\}$, which are independent of $r, s = 1, \ldots, S$, it follows that $n^{1/2}[\widehat{\theta}_n - S^{-1}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta^0)]$ is asymptotically normal with the variance-covariance matrix

$$[-J^{-1}, \frac{\widetilde{J}^{-1}}{S}, \ldots, \frac{\widetilde{J}^{-1}}{S}]\begin{bmatrix} \Sigma & \widetilde{K}_{0s} & \widetilde{K}_{0s} & \cdots & \widetilde{K}_{0s} \\ \widetilde{K}_{0s} & \widetilde{\Sigma} & \widetilde{K}_{rs} & \cdots & \widetilde{K}_{rs} \\ \widetilde{K}_{0s} & \widetilde{K}_{rs} & \widetilde{\Sigma} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \widetilde{K}_{rs} \\ \widetilde{K}_{0s} & \widetilde{K}_{rs} & \cdots & \widetilde{K}_{rs} & \widetilde{\Sigma} \end{bmatrix}\begin{bmatrix} -J^{-1} \\ \widetilde{J}^{-1}/S \\ \vdots \\ \widetilde{J}^{-1}/S \end{bmatrix}.$$

This matrix can be rewritten as

$$var(n^{1/2}\{\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta^0)\}) = J^{-1}\Sigma J^{-1} + \frac{1}{S}\widetilde{J}^{-1}\widetilde{\Sigma}\widetilde{J}^{-1} + (1 - \frac{1}{S})\widetilde{J}^{-1}\widetilde{K}_{rs}\widetilde{J}^{-1} - 2J^{-1}\widetilde{K}_{0s}\widetilde{J}^{-1} \tag{67}$$

where

$$\begin{aligned}
\widetilde{K}_{rs} &= E[\varphi_\tau(\widetilde{y}_i^r(\beta^0) - x_i^T\theta^0)x_i \cdot \varphi_\tau(\widetilde{y}_i^s(\beta^0) - x_i^T\theta^0)x_i] \\
&= E_x[E[\varphi_\tau(\widetilde{y}_i^r(\beta^0) - x_i^T\theta^0)|x_i] \cdot E[\varphi_\tau(\widetilde{y}_i^s(\beta^0) - x_i^T\theta^0)|x_i]x_i x_i^T) \\
&= E[\widetilde{F}_{\widetilde{y}(\beta^0)}(x_i^T\theta^0|x_i) - \tau)(\widetilde{F}_{\widetilde{y}(\beta^0)}(x_i^T\theta^0|x_i) - \tau)x_i x_i^T]
\end{aligned}$$

due to the independence of simulated errors $\{\widetilde{\varepsilon}_i^r\}_{i=1}^n$ and $\{\widetilde{\varepsilon}_i^s\}_{i=1}^n$, and consequently, of the simulated paths $\{\widetilde{y}_i^r(\beta^0)\}_{i=1}^n$ and $\{\widetilde{y}_i^s(\beta^0)\}_{i=1}^n$ (conditionally on x_i). Similarly, we can obtain $\widetilde{K}_{0s} = cov\{[F_y(x_i^T\theta^0|x_i) - \tau)x_i][(\widetilde{F}_{\widetilde{y}(\beta^0)}(x_i^T\theta^0|x_i) - \tau)x_i]\}$. By definition (10) of $\widetilde{F}_{\widetilde{y}(\beta^0)}$, $\widetilde{F}_{\widetilde{y}(\beta^0)}(x_i^T\theta^0|x_i) = F_y(x_i^T\theta^0|x_i)$ and thus $\widetilde{K}_{0s} = \widetilde{K}_{rs} = \widetilde{K}$. Combining (62) and (67) then yields the asymptotic distribution of $n^{1/2}(\widehat{\beta}_n^{III} - \beta^0)$, which is normal with variance

$$(D^T\Omega D)^{-1}D^T\Omega\left\{J^{-1}\Sigma J^{-1} + \frac{1}{S}\widetilde{J}^{-1}\widetilde{\Sigma}\widetilde{J}^{-1} + (1-\frac{1}{S})\widetilde{J}^{-1}\widetilde{K}\widetilde{J}^{-1} - 2J^{-1}\widetilde{K}\widetilde{J}^{-1}\right\}\Omega D(D^T\Omega D)^{-1}$$

(recall that $D = \partial b(\widetilde{F}(\beta^0), \beta^0)/\partial\beta^T$). Since D and Ω are full rank square matrices by Assumption A.6, the final expression follows from $(D^T\Omega D)^{-1} = D^{-1}\Omega^{-1}(D^T)^{-1}$. $\square$

Proof of Theorem 3: To establish the asymptotic equivalence result, we first need to prove that the feasible estimator $\widehat{\beta}_n^{FII}$ is consistent. The argument is the same as in the case of the infeasible estimator in Theorem 2 (the first two paragraphs of the proof) provided that we establish $\widehat{\theta}_n^s(\beta^0) \to b(\widetilde{F}(\beta^0), \beta^0) = b(F_y, \beta^0) = \theta^0$ and $\widehat{\theta}_n^s(\beta) \to b(\widetilde{F}(\beta), \beta) \neq \theta^0$ for $\beta \neq \beta^0$. This however follows from Lemma 11, stating that $\widehat{\theta}_n^s(\beta) - b(\widetilde{F}(\beta), \beta) = o_p(1)$ as $n \to \infty$ for any $\beta \in B$ and $s = 1, \ldots, S$. Theorem 2.7 of Newey and McFadden (1994) and the continuity of the one-to-one link function thus again implies that $\widehat{\beta}_n^{FII} \to \beta^0$ in probability as $n \to \infty$.

Next, taking the first-order derivative of the optimization problem (19) with respect to β, the first-order condition is obtained:

$$\left[\frac{1}{S}\sum_{s=1}^S \frac{\partial\widehat{\theta}_n^s(\widehat{\beta}_n^{FII})^T}{\partial\beta}\right]\Omega\left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^S \widehat{\theta}_n^s(\widehat{\beta}_n^{FII})\right] = 0. \tag{68}$$

Applying the Taylor expansion around β^0 to $\widehat{\theta}_n - \sum_{s=1}^S \widetilde{\theta}_n^s(\widehat{\beta}_n^{FII})/S$, we obtain for some linear combination ξ_n of β^0 and $\widehat{\beta}_n^{FII}$ that

$$\left[\frac{1}{S}\sum_{s=1}^S \frac{\partial\widehat{\theta}_n^s(\widehat{\beta}_n^{FII})^T}{\partial\beta}\right]\Omega\left[\left\{\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^S \widehat{\theta}_n^s(\beta^0)\right\} - \frac{1}{S}\sum_{s=1}^S \frac{\partial\widehat{\theta}_n^s(\xi_n)^T}{\partial\beta}\left\{\widehat{\beta}_n^{FII} - \beta^0\right\}\right] = 0. \tag{69}$$

We know from Lemma 11 that $\widehat{\theta}_n^s(\beta) \to b(\widetilde{F}(\beta),\beta)$. As the consistency of $\widehat{\beta}_n^{FII}$ and Corollary 9 imply that $\widetilde{F}(\beta)$ is continuous in $\beta \in U(\beta^0,\delta)$ and that $\|\widehat{F}(\widehat{\beta}_n^{FII}) - \widetilde{F}(\beta^0)\| = o_p(1)$ as $n \to \infty$, the continuity of the link function (Assumption A.6) implies that $\widehat{\theta}_n^s(\widehat{\beta}_n^{FII}) \to b(\widetilde{F}(\beta^0),\beta^0) = \theta^0$; similarly, $\partial\widehat{\theta}_n^s(\widehat{\beta}_n^{III})/\partial\beta^T \to \partial b(\widetilde{F}(\beta^0),\beta^0)/\partial\beta^T$, and $\partial\widehat{\theta}_n^s(\xi_n)/\partial\beta^T \to \partial b(\widetilde{F}(\beta^0),\beta^0)/\partial\beta^T$ in probability for $n\to\infty$.

Consequently, we have due to the full-rank Assumption A.6 that

$$\begin{aligned} n^{1/2}(\widehat{\beta}_n^{FII} - \beta^0) &= \left(\frac{\partial b^T}{\partial\beta}(\widetilde{F}(\beta^0),\beta^0)\Omega\frac{\partial b}{\partial\beta^T}(\widetilde{F}(\beta^0),\beta^0)\right)^{-1}\frac{\partial b^T}{\partial\beta}(\widetilde{F}(\beta^0),\beta^0)\Omega \\ &\times\ n^{1/2}\left[\widehat{\theta}_n - \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}_n^s(\beta^0)\right] + o_p(1). \end{aligned} \tag{70}$$

Recall that D denotes $\partial b(\widetilde{F}(\beta^0),\beta^0)/\partial\beta^T$. Subtracting (62) from (70) yields

$$n^{1/2}(\widehat{\beta}_n^{III} - \widehat{\beta}_n^{FII}) = (D^T\Omega D)^{-1}D^T\Omega\cdot n^{1/2}\left[\frac{1}{S}\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta^0) - \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}_n^s(\beta^0)\right] + o_p(1). \tag{71}$$

To prove the theorem, we have to show that the difference $\sum_{s=1}^{S}\widetilde{\theta}_n^s(\beta^0)/S - \sum_{s=1}^{S}\widehat{\theta}_n^s(\beta^0)/S$ is asymptotically negligible in probability. The estimates $\widetilde{\theta}_n^s(\beta^0)$ and $\widehat{\theta}_n^s(\beta^0)$ are obtained for data $y_i = \max\{x_i^T\beta + \varepsilon_i^s, 0\}$, where $\varepsilon_i^s \sim N(\mu_\tau\sigma(x_i), \sigma(x_i))$ and $\sigma(x_i)$ represent the conditional variance functions $\widetilde{\sigma}(x_i;\beta^0)$ and $\widehat{\sigma}_n(x_i;\beta^0)$, respectively. Thus, $\widetilde{\varepsilon}_i^s(\beta^0)|x_i$ follows a normal distribution with mean $\mu_\tau\widetilde{\sigma}(x_i;\beta^0)$ and variance $\widetilde{\sigma}(x_i;\beta^0)$ and $(x_i,\widetilde{\varepsilon}_i^s(\beta^0))$ follows their joint distribution P_0. On the other hand, $\widehat{\varepsilon}_i^s(\beta^0)|x_i$ is characterized a different conditional distribution, which is a normal distribution with mean $\mu_\tau\widehat{\sigma}_n(x_i;\beta^0)$ and variance $\widehat{\sigma}_n(x_i;\beta^0)$; the joint distribution of $(x_i,\widehat{\varepsilon}_i^s(\beta^0))$ is denoted P_n (it has the same marginal distribution of x_i as P_0).

We have already established that both $\widetilde{\theta}_n^s(\beta^0)$ and $\widehat{\theta}_n^s(\beta^0)$ are consistent estimators of θ^0. As the variance functions at β^0 are bounded (see the introduction of Appendix B), censored data simulated from P_0 and P_n satisfy assumptions of Theorem 5 if condition (28) is verified. As all terms of $\{r(y_i,x_i,\theta') - r(y_i,x_i,\theta'')\}^2$ in condition (28) with their expectations varying with P_n have the form $C_2(\theta',\theta'',\theta^0)\rho_\tau(y - x^T\theta')\rho_\tau(y - x^T\theta'')$, $C_1(\theta',\theta'',\theta^0)\rho_\tau(y - x^T\theta)I(y \le x^T\theta^0)$, or $C_0(\theta',\theta'',\theta^0)I(y \le x^T\theta^0)I(y \le x^T\theta^0)$ for some deterministic functions C_0, C_1, C_2 and some $\theta',\theta'' \in U(\theta^0,\delta)$, Lemma 10 implies the validity of condition (28). Consequently, we can write using the asymptotic linearity result of Theorem 5 for any $s = 1, \dots, S$

$$\begin{aligned} n^{1/2}\left[\widetilde{\theta}_n^s(\beta^0) - \widehat{\theta}_n^s(\beta^0)\right] &= -\widetilde{J}^{-1} G_{n,P_0}[\varphi_\tau(\max\{0, x_i^T\beta^0 + \varepsilon_i^s\} - x_i^T\theta^0)x_i] \\ &+ \ \widetilde{J}^{-1} G_{n,P_n}[\varphi_\tau(\max\{0, x_i^T\beta^0 + \varepsilon_i^s\} - x_i^T\theta^0)x_i] + o_p(1) \end{aligned}$$

as $n \to \infty$. Let us denote $g(x_i, \varepsilon_i^s; \beta, \theta) = \varphi_\tau(\max\{0, \varepsilon_i^s + x_i^T\beta\} - x_i^T\theta)x_i$. Then

$$n^{1/2}\left[\widetilde{\theta}_n^s(\beta^0) - \widehat{\theta}_n^s(\beta^0)\right] = \widetilde{J}^{-1}\left\{G_{n,P_n}[g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] - G_{n,P_0}[g(x_i, \varepsilon_i^s; \beta^0, \theta^0)]\right\} + o_p(1).$$

To verify the claim $n^{1/2}\left[\widetilde{\theta}_n^s(\beta^0) - \widehat{\theta}_n^s(\beta^0)\right] = o_p(1)$, we have to prove that $G_{n,P_n}[g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] - G_{n,P_0}[g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] = (G_{n,P_n} - G_{n,P_0})[g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] \xrightarrow{P} 0$ as $n \to \infty$. We only have to show that $(G_{n,P_n} - G_{n,P_0})[g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] \to 0$ in distribution since this is equivalent to the convergence to 0 in probability (cf., the proof of Van der Vaart, 2000, Lemma 19.24). This result follows directly from Van der Vaart and Wellner (1996, Theorem 2.8.9) once we verify their assumptions:

1. The class of functions $\mathcal{G} = \{g(x, \varepsilon; \beta, \theta) : \beta \in U(\beta^0, \delta), \theta \in U(\theta^0, \delta)\}, \delta > 0$, should satisfy the uniform entropy condition. Since $g(x, \varepsilon; \beta, \theta) = [\tau - I(\max\{0, x^T\beta + \varepsilon\} - x^T\theta \le 0)]x_i$, the class $\mathcal{G}$ forms a VC class by Van der Vaart and Wellner (1996, Lemma 2.6.18) and it thus satisfies the uniform entropy condition by Van der Vaart and Wellner (1996, Theorem 2.6.7).
2. Next, $P_n\bar{g}$ has to be bounded for an envelope function $\bar{g}$ of $\mathcal{G}$ and has to satisfy the Lindenberg condition $\limsup_{n\to\infty} P_n\bar{g}^2\{\bar{g} \ge \epsilon\sqrt{n}\} = 0$ for every $\epsilon > 0$. As the functions in $\mathcal{G}$ are bounded by a constant (see Assumption A.3), this requirement is also satisfied.
3. Finally, it has to hold that $\sup_{g,g'\in\mathcal{G}} |E_{P_n}(g - g')^2 - E_{P_0}(g - g')^2| \to 0$ as $n \to \infty$, where $E_P(g - g')^2 = E_P[I(\max\{0, x_i^T\beta + \varepsilon_i\} - x_i^T\theta \le 0) - I(\max\{0, x_i^T\beta' + \varepsilon_i\} - x_i^T\theta' \le 0)]^2 x_i x_i^T$, $(x_i, \varepsilon_i) \sim P$, and (β, θ) and (β', θ') correspond to functions g and g', respectively. This however directly follows from Lemma 10 as x_i has a compact support by Assumption A.3.

Consequently, $n^{1/2}\left[\widetilde{\theta}_n^s(\beta^0) - \widehat{\theta}_n^s(\beta^0)\right] = o_p(1)$ for any $s = 1, \dots, S$ and it follows from equation (71) that $n^{1/2}(\widehat{\beta}_n^{III} - \widehat{\beta}_n^{FII}) = (D'\Omega D)^{-1}D'\Omega n^{1/2}o_p(1) = o_p(1)$. □

[1] Arabmazar, A., and P. Schmidt, 1982, An investigation of the robustness of the Tobit estimator to non-normality. Econometrica 50(4), 1055–1063.

[2] Angrist, J., V. Chernozhukov, and I. Fernandez-Val, 2006, Quantile regression under

misspecification, with an application to the US wage structure. Econometrica 74(2), 539–563.

[3] Buchinsky, M., and J. Hahn, 1998, An alternative estimator for the quantile regression model. Econometrica 66, 653–672.

[4] Campbell, J., and B. E. Honore, 1993, Median unbiasedness of estimators of panel data censored regression models. Econometric Theory 9(3), 499–503.

[5] Chernozhukov, V., and H. Hong, 2002, Three-step censored quantile regression and extramarital affairs. Journal of the American Statistical Association 97, 872–882.

[6] Čížek, P., 2006, Least trimmed squares in nonlinear regression under dependence. Journal of Statistical Planning and Inference 136, 3967–3988.

[7] Davidson, J., 1994, Stochastic limit theory. Oxford University Press, Oxford.

[8] Fahr, R., 2004, Loafing or learning? – the demand for informal education. European Economic Review 49, 75–98.

[9] Fitzenberger, B., 1997a, A guide to censored quantile regressions, in: G. S. Maddala and C. R. Rao (Eds.), Handbook of Statistics, Vol. 15: Robust Inference. North-Holland, Amsterdam, pp. 405–437.

[10] Fitzenberger, B., 1997b, Computational aspects of censored quantile regression, in: Y. Dodge (Ed.), Proceedings of the Third International Conference on Statistical Data Analysis based on the L1-Norm and Related Methods, Vol. 31. Hayword, California, pp. 171–186.

[11] Gouriéroux, C., A. Monfort, and E. Renault, 1993, Indirect inference. Journal of Applied Econometrics 8, 85–118.

[12] Gouriéroux, C., E. Renault, and N. Touzi, 2000, Calibration by simulation for small sample bias correction, in: R. S. Mariano, T. Schuermann, and M. Weeks (Eds.), Simulation-Based Inference in Econometrics: Methods and Applications. Cambridge University Press, Cambridge, pp. 328-358.

[13] Gouriéroux, C., P. C. B. Phillips, and J. Yu, 2010, Indirect inference for dynamic panel models. Journal of Econometrics 157, 68–77.

[14] Hall, P., J. S. Racine, and Q. Li, 2004, Cross-validation and the estimation of conditional probability densities. Journal of the American Statistical Association 99, 1015–1026.

[15] Hahn, J., 1995, Bootstrapping quantile regression estimators. Econometric Theory 11, 105–121.

[16] Honore, B. E., 1992, Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. Econometrica 60(3), 533–565.

[17] Honore, B. E., and J. L. Powell, 1994, Pairwise difference estimators for censored and truncated regression models. Journal of Econometrics 64, 241–278.

[18] Honore, B., S. Khan, and J. L. Powell, 2002, Quantile regression under random censoring. Journal of Econometrics 109, 67–105.

[19] Horowitz, J. L., 1986, A distribution-free least squares estimator for linear censored regression models. Journal of Econometrics 32, 59–84.

[20] Khan, S., and J. L. Powell, 2001, Two-step estimation of semiparametric censored regression models. Journal of Econometrics 103, 73–110.

[21] Li, Q., and J. S. Racine, 2008, Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. Journal of Business and Economic Statistics 26, 423–434.

[22] Manski, C. F., 1975, Maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3, 205–228.

[23] Melenberg, B., and A. van Soest, A., 1996, Parametric and semi-parametric modeling of vacation expenditures. Journal of Applied Econometrics 11(1), 59–76.

[24] Moon, C.-G., 1989, A Monte Carlo comparison of semiparametric Tobit estimators. Journal of Applied Econometrics 4, 361–382.

[25] Newey, W. K., and D. McFadden, 1994, Estimation and hypothesis testing in large samples, in: R. F. Engle and D. McFadden (Eds.), Handbook of Econometrics, Vol. 4. North-Holland, Amsterdam.

[26] Paarsch, H. J., 1984, A Monte Carlo comparison of estimators for censored regression models. Journal of Econometrics 24, 197–213.

[27] Pakes, A., and D. Pollard, D., 1989, Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1058.

[28] Portnoy, S., 2003, Censored regression quantiles. Journal of the American Statistical Association 98, 1001–1012.

[29] Powell, J. L., 1984, Least absolute deviations estimator for the censored regression model. Journal of Econometrics 25, 303–325.

[30] Powell, J. L., 1986a, Censored regression quantiles. Journal of Econometrics 32, 143–155.

[31] Powell, J. L., 1986b, Symmetrically trimmed least squares estimation of Tobit models. Econometrica 54, 1435–1460.

[32] Racine, J. S., and Q. Li, 2004, Nonparametric estimation of regression functions with both categorical and continuous data. Journal of Econometrics 119, 99–130.

[33] Tang, Y., H. J. Wang, X. He, and Z. Zhu, 2011, An informative subset-based estimator for censored quantile regression. TEST, 1–21.

[34] Van der Vaart, A. W., 2000, Asymptotic Statistics. Cambridge University Press, Cambridge.

[35] Van der Vaart, A. W., and J. Wellner, 1996, Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, New York.

[36] Xia, Y., H. Tong, W. K. Li, L.-X. Zhu, 2009, An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society, Series B, 64(3), 363–410.